



REFERENCE

NIST  
PUBLICATIONS

# A Comparative Study of CMM Flat-Surface Inspection Plans

**J. P. Hsu****T. W. Hsu**

Department of Mechanical and  
Industrial Engineering  
University of Texas at El Paso

**James J. Filliben**

U.S. DEPARTMENT OF COMMERCE  
Technology Administration  
National Institute of Standards  
and Technology  
Statistical Engineering Division  
Gaithersburg, MD 20899

**Theodore H. Hopp**

U.S. DEPARTMENT OF COMMERCE  
Technology Administration  
National Institute of Standards  
and Technology  
Factory Automation Systems Division  
Gaithersburg, Md 20899

QC

100

.U56

4897

1992

**NIST**







## **ABSTRACT**

This document reports the findings from a collaborative study by the authors of inspection sampling plans as related to coordinate measuring machines (CMM's). The goal of the study was to compare various inspection plans, and, in particular, to make recommendations on minimal sample sizes needed to achieve pre-specified engineering tolerances on a CMM. The investigation was limited to the study of plans to evaluate the parallelism of planar surface.

The basis of the study was an evaluation of the performance of different sampling plans in assessing surface parallelism for several physical artifacts. This report describes how data were acquired, the sampling plans that were studied, the data analysis performed, and the results of the analysis.



# TABLE OF CONTENTS

1 INTRODUCTION . . . . .	1
Coordinate Measurement Machines . . . . .	1
Experimental Goals . . . . .	1
Our General Approach to the Problem . . . . .	1
2 DATA ACQUISITION . . . . .	2
3 THE DATA ANALYSIS . . . . .	3
Anomalies Analysis . . . . .	4
Sampling Plan/Sample Size Determination . . . . .	5
Sampling Plans . . . . .	5
Sample Size Requirements . . . . .	6
Procedural Details . . . . .	7
Graphical Analysis . . . . .	7
Bias Correction for Sample Ranges . . . . .	10
High-Frequency Surface Characterization . . . . .	11
4 CONCLUSIONS . . . . .	12
5 ACKNOWLEDGMENTS . . . . .	14
6 REFERENCES . . . . .	14



# 1 INTRODUCTION

## 1.1 Coordinate Measurement Machines

A typical coordinate measurement machine (CMM) consists of a gage head with a mounted probe and stylus, and a 3-axis numerical readout. CMM's are often run under computer control, are programmable, and have automated data acquisition systems. A typical CMM is capable of doing discrete, point-to-point measurements at up to 1- $\mu$ m resolution in 3 axes. Because of their flexibility in measuring any rigid surface, CMM's have come to be considered as the *de facto* precision "universal gage," and have become a primary on-line and off-line tool in modern complex-geometry manufacturing. CMM part data is frequently processed to create "substitute geometries" which may then be subsequently compared to nominal/ideal geometries to ascertain whether a part is in tolerance. However, although the output of CMM's is heavily relied on and used, the performance sensitivity and bias of CMM's on "real world" surfaces is neither well-appreciated nor well-understood.

## 1.2 Experimental Goals

An important engineering objective in parts inspection is to estimate parameters of a "best-fit" geometry to the part (e.g., center point and radius of the "best-fit" circle to a slice from a cylindrical part). This is frequently accomplished by the CMM taking "enough" (4 to 8) points at "selected" locations on the part. These points are then statistically processed so as to arrive at the desired geometry parameter estimates. (See Feng and Hopp [1] for a survey of fitting algorithms used on CMM's.) The question of "how many" points and "where are they sampled" are thus the two key questions at the heart of CMM inspection sampling plans. Too many points result in wasted machine time; too few points result in poor geometry parameter estimates.

The primary goal of this study was to determine the minimal required sample size needed in CMM inspection plans so as to have geometry part estimates within achieve pre-specified engineering tolerances. To come up with conclusions that have industrial relevance, the performance of the CMM in relation to a variety of "real world" surfaces was a key part of the investigation. This study focused on the parallelism of planar surfaces to given datum planes. However, we believe that our general approach can be applied to more complex surface geometries and characteristics.

## 1.3 Our General Approach to the Problem

Three approaches may be taken to research CMM inspection algorithms and sampling techniques:

1. theoretical analysis without experimentation,
2. experimental studies with simulated data, and
3. experimental studies with real, measured data.

Our focus has been on the third approach. We believe that this is a major improvement over the commonly used simulated sinusoidal surfaces. Having such data sets available has already served as the basis for the minimal sample size determination for the completely randomized and randomized block sampling plans, and will serve equally-well in the future as the basis for evaluating and comparing other sampling plans that come under consideration. Our general approach can be broken down into the following steps:

1. Select and identify artifacts representing a range of surface characteristics.
2. Densely measure the selected artifacts with a CMM. This will constitute our "population".
3. Compute reference values for key geometric parameters based on the population data.
4. Simulate various sampling strategies with regard to the number and placement of points by selecting subsets of the population.
5. Estimate the same key parameters from the subsets.
6. Evaluate the various sampling plans by graphical and quantitative comparison of parameter estimates in relation to the population reference values.

## 2 DATA ACQUISITION

All the data used in the analysis was acquired using a Sheffield Cordax\* CMM available in the Automated Manufacturing Research Facility (AMRF) of NIST. Data points were gathered with a 0.8-mm ball mounted on a Renishaw touch-trigger probe. The CMM was operated according to commonly accepted practice with the following motion parameters: 7.6 mm/s touch speed, 125 mm/s traverse speed, and 0.2 mm stand-off distance. In detail, we did the following.

1. A program was written in the C programming language for an IBM PC computer to directly control the data collection process on the Sheffield Cordax CMM.

---

\*Certain trade names and company products are mentioned in the text or identified in an illustration in order to adequately specify the experimental procedure and equipment used. In no case does such identification imply recommendation or endorsement by the National Institute of Standards and Technology, nor does it imply that the products are necessarily the best available for the purpose.

2. A glass optical flat was measured using a light box. It was flat within  $0.025 \mu\text{m}$ . We also measured the profile of a line on this optical flat using a CMM. For the line, four hundred points were gathered at equal intervals over approximately 5 cm.
3. We obtained a National Aerospace Standard artifact (NAS 979) and measured two line profiles. These were located on the part as shown by the dotted lines in Figure 2. Again, four hundred points were gathered at equal intervals over 5 cm.
4. We obtained an extensive set of planar artifacts made by General Electric as standard roughness specimens. (These specimens approximate various machined surfaces and represent different degrees of roughness commonly encountered in surface processing. The description of this set of specimens is shown in Figure 3.) We measured a line profile on each of the General Electric (GE) Artifacts A, B, C, D, E, F and G, and two lines on GE Artifact H. As before, four hundred equally-spaced points were collected over a distance of 5 cm.

In all, twelve populations of data were acquired. This includes one line trace on each of the GE artifacts A through G, two traces on GE artifact H, one trace on the optical flat, and two traces on the NAS test part. Throughout the rest of this paper, each of the twelve data sets will be referred to as an artifact. The reader should bear in mind, however, that only ten physically distinct parts were used.

### 3 THE DATA ANALYSIS

The data analysis proceeded in four stages. We briefly describe each stage here, and treat it in detail later. Most of the analysis was carried out using the Dataplot® statistical analysis and graphics package developed at NIST. (Dataplot is described in Filliben [2].)

1. Anomalies Analysis. Was there anything in the data that suggested anomalous or unexpected behavior in the artifact surfaces, in the CMM measuring process, or in interactions between the CMM and the surface? What Exploratory Data Analysis\*\* techniques serve well in detecting such anomalies?
2. Sampling Plan/Sample Size Determination. What sampling plan is better? For a given sampling plan (e.g., completely randomized design), what is the minimal sample size required so that with high confidence (e.g., 90%), a certain engineering objective is achieved? Does the sample size vary across different surfaces/machining types? How

---

\*\*Exploratory Data Analysis is that area of statistics whose goal is to provide diagnostic insight into underlying structure. This is in contrast to other branches of statistics dealing with data summarization and model confirmation. See Tukey [3].

do such sample sizes compare with current industrial practices? How can differences among plans be quantified?

3. Bias Correction. Are the commonly-used performance statistics (e.g., the sample range) unbiased or biased? (That is, are the statistics distributionally centered on the reference value?) If biased, what correction factors are needed to correct the bias? Are the correction factors the same across all surfaces/machine types?
4. Surface Characterization. Of the ten physical artifacts, do any statistical tools (spectral plots, lowess, etc.) serve to provide additional methodologies to characterize the ten surfaces?

### 3.1 Anomalies Analysis

Anomalies analysis is intended to assess the quality of the data. The reader is directed to Figures 4, 5, and 6. Figure 4 shows the four-hundred-point surface profile of the optical flat artifact. The flatness of this artifact was independently established as  $0.025\ \mu\text{m}$ . However, the range in Figure 4 is considerably larger. Twelve outliers, with a range of approximately  $8\ \mu\text{m}$ , can be seen. Even with such outliers set aside, the range (max - min) of the data is approximately  $3\ \mu\text{m}$ . Furthermore, the range of any local subset of data (say ten successive observations) is about  $1\ \mu\text{m}$ . Ideally, these plots should all consist of a single horizontal line to reflect the ultra-smooth nature of the optical flat. In fact, however, we see evidence of four separate sources of variability:

1. anomalies/outliers ( $8\ \mu\text{m}$ );
2. CMM intrinsic resolution ( $1\ \mu\text{m}$ );
3. setup/alignment errors ( $1\ \mu\text{m}$ );
4. other measurement errors ( $1\ \mu\text{m}$ ).

The anomalies/outliers did not reproduce on a different CMM at the University of Texas, El Paso. The most likely explanation of the anomalies is that the combination of standoff distance and touching speed set on the NIST CMM did not allow enough time for the ram oscillations on the CMM to damp out before the surface was contacted. The intrinsic resolution of the CMM can be seen in the  $1\text{-}\mu\text{m}$  high-frequency variation. The setup/alignment error appears as a linear drift in the data values. The cause of the remaining  $1\text{-}\mu\text{m}$  errors is unknown, but may be due to unidentified CMM errors or to environmental variability.

Figures 5 and 6 show surface profiles for the GE Artifact A; Figure 5 shows the first 50 (out of 400) points; Figure 6 shows the first 100 (out of 400) points. These subset blowups show similar interesting behavior as first seen in Figure 4. Note in Figures 5 and (especially) 6 how there is an apparent  $3\ \mu\text{m}$  gap in the data—a heavy trace of points above, and a thinner

parallel trace of points 3  $\mu\text{m}$  below. The physical explanation for this is unknown at this time. Note also in Figures 5 and 6 the discrete nature of the points, that data points occur in strata at -1  $\mu\text{m}$ , at -2  $\mu\text{m}$ , etc. As with Figure 5, this is an empirical verification of the rated 1- $\mu\text{m}$  resolution of the Sheffield CMM.

### **3.2 Sampling Plan/Sample Size Determination**

The next stage of analysis is to compare the performance of different sampling plans and to establish sample size requirements for each plan. A sampling plan is a strategy for selecting which points are to be measured. Sample size is the number of points to be measured according to a given plan. It is well-known that some sampling plans are better (in terms of reducing bias and variance of resulting statistics) than others. One of our objectives was to quantify the performance of various plans for the twelve data populations. We also felt that the sample size needed to achieve a particular engineering objective would be strongly influenced by the characteristics of the surfaces being inspected. So a second goal of the analysis was to examine the relationship between sampling plan, surface, and sample size.

In this section, we first discuss the sampling plans that were considered. Next, we precisely frame the question of establishing sample size requirements. We then present the details of our analysis procedures. Fourth, we present the analysis, based on graphical techniques, in which we study sample size requirements, their relationship to surface characteristics, and relative efficiency of different sampling plans.

#### **3.2.1 Sampling Plans**

Each artifact is represented by a population of four hundred points. A sampling plan is a strategy for selecting a particular number of points from among the four hundred. Our approach to evaluating various sampling plans is to examine how simulating measurement processes behave when measurement data are drawn from the population according to each plan. We considered two widely used plans:

1. completely randomized sampling;
2. randomized block (stratified) sampling.

The core question is:

Out of the population of four hundred points, how representative (of the population) is a sample of, say, twenty points collected via a completely randomized plan? Would it be more (or less) representative than a set of twenty points collected via a randomized block plan?

Of the two plans, the most basic is completely randomized sampling. In this plan, sample points are randomly drawn from the population, with each point in the population having an equal chance of selection. We have been able to arrive at specific conclusions about minimal sample sizes for this type of sampling plan. The results are interesting, somewhat surprising,

and as usual are suggestive of further research. A good portion of the remainder of this report summarizes our findings with respect to completely randomized sampling plans.

The second sampling plan considered is the randomized block design. In general terms, this plan is defined by partitioning a factor of the experiment into blocks and randomly sampling within each block. In our case, the blocking factor is the position along the 5 cm sampling line on the part surface. We examined one particular randomized block plan: a four-block plan with block size set at a quarter of the total extent of the artifact, or a span of one hundred points.

### 3.2.2 Sample Size Requirements

The primary question which must be addressed for our CMM analysis is as follows:

For a given artifact, and for a given sampling plan, what is the minimal number of observations,  $n^*$ , that one needs to collect so that some given engineering criteria is attained?

The remainder of this report answers that question for the completely randomized sampling scheme and for the randomized block design (four blocks).

The first related question that needs clarification is what is the "given engineering criteria" that needs to be attained? We used a very simple criterion: the parallelism of a line to the plane established by the bottom of the artifact. Consider the engineering case whereby two planar surfaces are to be mated face-to-face. For many applications, it is extremely important that each surface be parallel to a reference coordinate plane within a specified tolerance, or else leakage, noise, or some other manifestation of sub-par mechanical performance will result. For a given surface, how does one measure such parallelism? A cross-section of two such mating surfaces would be represented by two mating traces. Each such trace would ideally be linear and parallel to the datum reference plane. More commonly it is not.

One measure of parallelism deviation of a trace is the range statistic. The range is the difference between the high and low points on the trace. Ideally the range should be small and within the parallelism tolerance. The range precisely captures the notion of parallelism for the sample values. The point is, of course, that the sample range may not accurately represent the population range.<sup>\*\*\*</sup>

From an engineering point of view, a complete dense sampling of a line on the surface, would yield a straightforward value for the range—we refer to that idealized target range value as the population range  $R_{pop}$ . Unfortunately, such a complete dense sampling is prohibitively time-consuming and expensive. It thus becomes an important practical engineering problem to be able to estimate the range  $R_{pop}$  while only having collected a small sample of points from the surface. In this light, the core engineering question thus becomes:

---

<sup>\*\*\*</sup>For engineering applications in which flatness is important, other statistics which are insensitive to orientation/tilt should be utilized (e.g., the width of the tightest parallel plane envelope).

How many observations  $n^*$  does one need to assure with high probability (e.g., 90%) that the range  $R$  computed from a sample is within some relative tolerance percent (e.g., 20%) of the true, unknown population range  $R_{pop}$ ?

The answer to this question is not yet fully understood. Also, how the needed sample size  $n^*$  is related to surface roughness parameters—and ultimately, to surface/material machining operations—is also unknown. The remainder of this section provides (for completely randomized designs) initial answers to these questions.

### 3.2.3 Procedural Details

We proceeded as follows: On each of the twelve artifacts (ten physical artifacts plus an additional replication on the GE Artifact H and another on the NAS standard), a dense set of 400 points was collected along a 5-cm straight line. This set of 400 points serves as our population for the given artifact. The procuring and measuring of these 12 artifacts was a somewhat time-consuming activity, but well worth it because these 12 population sets now serve as "standard reference data sets" for 10 different "real world" surfaces.

For a given artifact, a population of 400 points was thus obtained, and so the desired population range  $R_{pop}$  (largest point minus smallest point) was computed. We thus have (for each artifact) a statistic  $R_{pop}$  which is a measure of the maximal peak-to-valley discrepancy across the entire population.  $R_{pop}$  hence serves as the reference value by which the corresponding statistic in samples of size 4,8,12,... will be compared. A simulation was then carried out for each artifact by drawing random samples of various sizes from each population of 400. For each sample size 4,8,12,...,40 and for each of the 12 artifacts, 100 such random samples were drawn. For each of the 100 samples, the sample range (max - min of the sample data) was computed.

For a given artifact, a sample size would be considered "good" if "most" of the set of 100 sample ranges was sufficiently "close" to the population range  $R_{pop}$ . How much is "most"? This is an engineering question. In our analysis, "most" was set at 90%—we wanted 90% of the sample ranges to be "close" to the reference population range  $R_{pop}$ . How close is "close"? This also is an engineering question which depends on the seriousness of the engineering application (e.g., aircraft/aerospace applications dictate a higher degree of "closeness" than would household applications) and on the criticality of the engineering component. For analysis purposes, we chose a tolerance factor of 20%—in other words, we will consider a sample range  $R$  to be "close enough" to the population range  $R_{pop}$ , if  $R$  is within 20% of  $R_{pop}$ —that is, if the  $R/R_{pop} > 80\%$ . We use this 80% cutoff value for the remainder of this report, although other cutoff values could also have been utilized and analyzed in a similar fashion.

### 3.2.4 Graphical Analysis

Figures 7-a through 7- $\ell$  graphically summarize the results of our simulation study for the twelve artifacts. Figure 7-a, for example, shows the results from the GE Artifact A analysis. Figure 7-a consists of 6 plots—each plot has the sample size  $m$  on the horizontal axis.

The first (upper left) of the 6 plots in Figure 7-a shows the total distributional spread of the 100  $R/R_{pop}$  value for each sample size. The range of the  $R/R_{pop}$  ratio is 0 to 1; a ratio near zero indicates the undesirable situation whereby the sample range is much removed from the

population range; a ratio near unity indicates the ideal situation whereby the sample range is approximately the same as the population range. Unfortunately, it is noted in Figure 7-a that many of the smaller sample sizes  $n$  yield  $R/R_{pop}$  ratios that cover almost the entire spectrum from very good to very bad. It is not until the sample sizes are well above 20 that the sample ranges universally start approaching the population range.

Where does the critical crossover sample size  $n^*$  occur? If we apply our engineering criterion of  $R/R_{pop}$  being  $\geq 0.8$  and if we apply our confidence factor of 90%, what is the required minimal sample size  $n$ ? To be specific:

What is the minimal sample size  $n$  such that 90% of the computed sample ranges are within 20% of the population range?

To answer this question, consider the other five plots in Figure 7-a. Plot 2 (the upper middle plot) is a box plot of the data in Figure 7-a. For the one hundred ranges at a given sample size  $n$ , the box plot displays the:

1. largest value (top of the upper bar),
2. estimated 75 percentile (bottom of the upper bar),
3. estimated 50 percentile (= median)
4. estimated 25 percentile (top of the lower bar), and
5. smallest value (bottom of the lower bar).

Noting the cutoff ratio value of 0.8 on the vertical axis, and noting that the median value (the  $x$ ) for sample size  $n = 10$  is at  $R/R_{pop} = 0.8$ , then we would conclude (but with only 50% confidence) that  $n = 10$  is an acceptably large enough sample size. This conclusion could also be read from plot 3 (upper right), which has all of the box plot information removed except for the median values.

Plot 4 (lower left) in Figure 7-a was also derived from the box plot—it captures the top of the lower bar (the estimated 25% point) values of the box plot. Again referring to the  $R/R_{pop}=0.8$  cutoff value, we deduce from plot 4 that a sample size of approximately  $n = 16$  will assure (but with only 75% confidence) that the sample ranges will be "close enough".

For 90% confidence, we refer the reader to plot 5 (lower middle) of Figure 7-a which plots the first decile (the 10% point) of the data at each  $n$ . From this plot and using the same 0.8 cutoff value, we deduce that we must have a sample size of about 26 before we can be 90% confident that  $R/R_{pop} \geq 0.8$ . The value of  $n = 26$  is thus the desired  $n^*$  minimal sample size that we seek for the GE Artifact A. Although pessimistically large, it nevertheless is reflective of reality in sampling even from a smooth artifact such as GE Artifact A. Such tolerance-limit kind of information is extremely important to the engineer in realistically specifying inspection plans. It is seen from all five of the plots how grossly the sample range underestimates the population range for those sample sizes of 5 and smaller, commonly used in industry.

The lower right plot in Figure 7-a is a plot of the observed minima (the smallest of the 100 computed  $R/R_{pop}$  values). In a very simplistic sense, this may be considered as yielding information at the near-100% confidence level. We note that the desired sample size by which we are "100% sure" that the 100  $R/R_{pop}$  values exceeds 0.8 is well beyond the sample sizes (as dictated by industrial practice) included in our study.

Thus of the six plots in Figure 7-a, the most important plots are plot 2 (the box plot) and plot 5 (the decile plot). For providing a specific answer to our minimal sample size question at the 90% confidence level, plot 5 is to be used. The same six plots for the other eleven data sets are shown in Figures 7-b through 7-l.

Figures 7-a through 7-l demonstrate that the minimal sample size value  $n^*$  varies significantly across the artifacts. The GE Artifacts go from smooth (Artifact A, Figure 7-a) to rough (Artifact H, Figures 7-h and 7-i), the optical flat (Figure 7-j) is ultra-smooth, and the NAS standard (Figures 7-k and 7-l) is of moderate smoothness. Table 1 summarizes the minimal sample size  $n^*$  information as extracted from plots 3, 4, and 5 of Figures 7-a to 7-l.

**Table 1. Minimal Sample Size such that  $R/R_{pop} \geq 0.8$   
Completely Randomized Design**

<u>Artifact</u>	<u>50% Confidence</u>	<u>75% Confidence</u>	<u>90% Confidence</u>
A-4-15	10	16	26
B-8-28	40	68	65
C-16-56	22	40	48
D-32-118	8	14	24
E-63-220	10	22	32
F-125-455	12	24	36
G-250-875	16	28	44
H-1-500-1750	8	14	20
H-2-500-1750	8	14	20
Optical Flat	8	12	20
NAS-1	42	48	60
NAS-2	44	48	60

We see from this table, that (for 90% confidence) there are no artifacts in which the minimal  $n^*$  is smaller than 20, and in many cases, the minimal  $n^*$  in fact exceeds 30. It also appears that  $n^*$  is not closely correlated with surface roughness. (For instance, the roughest GE artifact has the same  $n^*$  as the optical flat.) This suggests that surface form deviations and tilt effects, rather than surface roughness, are the major contributors to the range statistic for each artifact.

These conclusions are supported by the graphical presentation in Figures 8-a to 8-c. Figure 8-a is the collection of all of the box plots from Figures 7-a to 7-l, but organized by artifact; similarly, Figure 8-b is the median plots across the 12 artifacts, and Figure 8-c is the first decile plots across the 12 artifacts. It is clear from these plots that  $n^*$  varies across the artifacts, but it is not monotonic with roughness.

The magnitude of  $n^*$  for the twelve data sets is also important. Even for only 50% confidence there are no cases in which fewer than eight points are required. This can be compared to the common industrial practice of sampling only a "handful" (five to ten points) of

points. Our results strongly suggest that common industrial practice routinely results (at least half the time) in errors of more than 20% in parallelism measurements.

A similar analysis can be carried out for randomized block sampling plans. (Refer to Section 3.2 for a description of the four-block, randomized block plan we used.) The analysis plots for randomized block plans are shown in Figures 9-a to 9-c. These figures show the same information for the randomized block plans that are shown Figures 8 for completely randomized designs. In general, for a given artifact, the traces in Figure 9 are the same as or slightly higher than the corresponding traces in Figure 8. This means of course that the required minimal sample sizes for randomized block designs are slightly smaller than those for completely randomized designs. This fact is graphically summarized in plots 10-a to 10-b in which the solid line is for completely randomized design and the dotted line is for randomized block design. Table 2 is a side-by-side comparison of the minimal sample size for each of the twelve artifacts and for each of the three confidence levels.

**Table 2. Minimal Sample Size such that  $R/R_{pop} \geq 0.8$   
Completely Randomized Design (CRD) & Randomized Block Design (RBD)**

<u>Artifact</u>	<u>50% Confidence (CRD/RBD)</u>	<u>75% Confidence (CRD/RBD)</u>	<u>90% Confidence (CRD/RBD)</u>
A-4-15	10/10	16/16	26/22
B-8-28	40/28	68/34	65/36
C-16-56	22/20	40/36	48/52
D-32-118	8/8	14/14	24/18
E-63-220	10/12	22/20	32/28
F-125-455	12/12	24/20	36/32
G-250-875	16/16	28/24	44/48
H-1-500-1750	8/8	14/12	20/16
H-2-500-1750	8/8	14/12	20/16
Optical Flat	8/7	12/12	20/16
NAS-1	42/40	48/48	60/60
NAS-2	44/40	48/50	60/60

The general conclusion from this table is that randomized block designs are more efficient (in terms of  $n^*$ ) to completely randomized designs.

### 3.3 Bias Correction for Sample Ranges

From the plots in Figures 7-a to 7-l, it is clear that the sample range is biased low relative to the population range. (All the  $R/R_{pop}$  traces are below unity, indicating that all the sample ranges are below the population range.) The question then arises as to whether and how this bias can be corrected. A simple solution to this may be read directly off plot 3 (the median plot) for each artifact. For example, for GE Artifact A, and for sample size 10, plot 3 of Figure 7-a

tells us that the 50% split of the 100 computed sample ratio was approximately 0.8. This then suggests that a  $1/0.8 = 1.25$  correction factor for the sample range  $R$  would shift the entire distribution up so that it is median-centered on the derived population size  $R_{pop}$ . This will result in half the adjusted ranges being above the population range and half will be below.

A summary of these median-bias correction factors is provided in Figure 11—for the GE Artifact A and for sample sizes 4 (4) 40, that is for sample size 4 to sample size 40 in steps of 4. Note that the correction factor is of the order of 1 to 2 and decreases as the sample size increases. Similar size tables may be constructed for the other artifacts and for the randomized block design. The correction factor for the randomized block design will generally be smaller than for completely randomized designs. When available, it is preferable to make use of such bias-correction factors, although caution must be exercised. Even though the resulting adjusted sample ranges are median-centered on the population range, the adjusted ranges still have considerable variability—especially for the smaller sample sizes.

### 3.4 High-Frequency Surface Characterization

A separate question to be addressed is as follows: Can we characterize artifacts by means of their high-frequency components? The artifacts are known to differ in smoothness from GE Artifact A (most smooth) to GE Artifact H (least smooth). As seen from Columns 3 and 4 of Figure 3, the machining process was such that the surface texture was imposed by varying (vertical) peak-to-valley differences. Thus GE Artifact A (whose full coded name is GE A-4-15) has an average absolute deviation from a center line of  $0.1 \mu\text{m}$  ( $4 \mu\text{in}$ ) and an average peak-to-valley difference of  $0.4 \mu\text{m}$  ( $15 \mu\text{in}$ ), while GE Artifact H (whose coded name is GE H-300-1750) has an average absolute deviation of  $7.7 \mu\text{m}$  ( $300 \mu\text{in}$ ) and an average peak-to-valley difference of  $45 \mu\text{m}$  ( $1750 \mu\text{in}$ ).

Can these artifacts be additionally characterized by not only vertical—but also horizontal—frequency structure? Was such frequency structure passively imposed by the machining process? To examine this, the reader is directed to Figures 12 to 18.

Figure 12 shows the surface profiles of GE Artifact A. A strong low frequency (spatial drift) component is apparent, but there is no evidence of high frequency structure. Figure 13 exhibits all 12 artifacts on a common vertical scale. The top 4 plots (GE Artifacts A through D) have differing low frequency structure, but appear quite similar otherwise. Plots 4 to 9 of GE Artifacts E, F, G, and two runs on GE Artifact H exhibit increasing vertical and horizontal roughness structure. Plot 10 shows the ultra-flat nature of the optical flat—it deserves a ranking above GE Artifact A. Plots 11 and 12 (the two runs on the NAS 979 Standard) are similar to one another and have a roughness appearance that would rank in the middle of the GE artifacts.

The corresponding low and high frequency components of the 12 artifacts are shown in Figures 14 and 15, respectively. The low frequency structure was derived by applying the lowess procedure (locally-weighted scatter-plot smoothing) to the raw surface profile. (The lowess procedure is a smoothing technique that is unaffected by data outliers. See Chambers, *et al.* [4] for details.) Comparing the original data in Figure 13 to the smoothed data in Figure 14, we see that the smoothing process has successfully captured such low-frequency structure.

The high frequency structure was derived by subtracting the smoothed values from the raw data. The results across the 12 artifacts are shown in Figure 15. Note that GE Artifacts A

through D appear quite similar. The (vertical) amplitude and the (horizontal) period of the traces increase monotonically with roughness for GE Artifacts E, F, G, and especially the 2 runs on GE Artifact H. The ultra-smooth nature of the optical flat is apparent. It also appears that the two NAS data sets are similar in high-frequency structure to GE Artifacts E and F.

It appears that, aside from GE Artifacts A through D, the high-frequency structure is well-correlated with surface roughness (as determined by independent means). We can speculate on why Artifacts A through D appear similar. Two explanations are possible. First, the raw data may not have had sufficient resolution to discriminate among the four artifacts. The other GE artifacts showed a correlation between surface roughness amplitude and spatial frequency. If this correlation holds for the smaller roughness of Artifacts A to D, the probe on the CMM may have been too large to be sensitive to detect the surface differences. A second explanation is that the smoothing procedure used to obtain the high-frequency structure may have removed whatever differences were present in the raw data for Artifacts A to D.

Figure 16 displays the lag plot ( $Y_i$  versus  $Y_{i-1}$ ) for each of the 12 artifacts. (The lag plot is a technique for displaying different kinds of autocorrelation structure in data. See Box *et al.* [5] for details.) In a surface where only small random perturbations from a base line exists as one proceeds from one point to the next, the resulting lag plots would be a random point cloud without structure. Conversely, any systematic structure in the lag plot would imply a serial relatedness in the artifact, in the measuring device, in the measuring process, etc. Such structure may itself be used as a "fingerprint" or "signature" of the machining process—to be used in addition to other such "fingerprints" such as profile, histogram, and spatial plots. Note in Figure 16, the distinct lag plot pattern exhibited by the 12 artifacts. The artifact closest to the random ideal would be GE Artifact E. Note the tilted lattice/stratification effect in many of the plots - especially the optical Flat, GE Artifact D, NAS-1, NAS-2, and GE Artifact B. As one might expect, high autocorrelation structure exists for the 2 runs on the relative rough GE Artifact H. The lattice structure in the lag plot for the optical flat is particularly interesting. Since the artifact is known to be ultra-smooth, all of the observed structure must be due to the measurement process. It seems reasonable that this structure would also appear in the plots for the other artifacts, but it does not. Thus, the "signature" of each process actually represents both the effects of the surface variability and the characteristics of the measurement process.

Figure 17 displays the histograms for the 12 artifacts. Again, the thrust is to apply analysis techniques in an attempt to extract discriminating underlying structure—which may ultimately be linked to the machining process. Note the bimodal nature of GE Artifact A to GE Artifact D, the skewness of GE Artifact F and GE Artifact G, and the outliers in the optical plat. The histogram is, however, orientation/tilt dependent, and so some of the distinguishing histogram features may change when the orientation/tilt is subtracted out.

Finally, the reader is referred to Figure 18 which is a spectral plot of the high frequency component (residuals after lowess) of the raw data. These plots are particularly interesting because they complement nicely the previous "fingerprints" for the 12 artifacts. A white noise process is indicated by a (relatively) flat spectrum. The optical flat comes closest to this ideal. The NAS-1 and NAS-2 artifacts are next closest—but both have slightly more high-frequency content. GE Artifacts A through D are relative flat, but all have a prominent spike—though not positioned in a monotonic (with respect to surface smoothness) fashion. GE Artifact E is near sinusoidal with a dominant high-frequency spectrum, while GE Artifact F has a dominant mid-

frequency spike—the machining process for GE Artifact E and GE Artifact F is yielding a surface which may be described as a single-frequency sinusoidal.

GE Artifacts F, H1, and H2 all have multiple dominating low frequency spikes—this indicates that the machining process yields a surface consisting of a few low-frequency sinusoidal components. Many machining processes consist of rotary operations. The intrinsic circular nature of the various processes is nicely captured by the spatial technique.

## 4 CONCLUSIONS

The findings of our research are:

1. For the 12 lines that we measured, we were able to measure and ascertain the hardware resolution and it agreed with the rated 1- $\mu\text{m}$  resolution of the CMM.
2. Even for the ultra-smooth (0.025  $\mu\text{m}$ ) optical flat, unexplained outliers of the order of 8  $\mu\text{m}$  and low frequency variation of the order of 2  $\mu\text{m}$  were detected. Anomalous 1- $\mu\text{m}$  transition jumps in certain smooth artifacts were detected. CMM machine dynamics, due to high probing speeds and small stand-off distance, is the probable cause of both the outliers and the smaller, anomalous jumps.
3. Even for simple line geometry and a simple performance criteria (the range), the completely randomized design yields (at the 90% level) pessimistically large minimal sample sizes  $n^*$  ( $\geq 20$ ) for all artifacts—thus the common engineering practice of a "handful" of samples is grossly inadequate.
4. For more complicated geometries (planes, circles, cylinders, cones, etc.) and for more complicated performance measures (slopes, volumes, radii of circumscribed circles, etc.), the required minimal sample size can be studied in a fashion similar to our methodology. We conjecture that for these other geometries the sample sizes currently used in industry are probably just as inadequate as we have found in the case of lines.
5. The linkage between minimal sample size and smoothness does not appear to be monotonic. This suggests that the minimal sample size,  $n^*$ , is as strongly influenced by form and orientation as by surface roughness.
6. The randomized block design using four blocks is superior to the completely randomized design in that the required minimal sample sizes are smaller.
7. Based on the median  $R/R_{\text{pop}}$  plots, correction factors are computable so as to remove bias in the simple ranges and thus allow sample ranges to be median-centered on the population range.

8. Several exploratory data analysis techniques were used to successfully discriminate among several of the artifacts. We believe that these techniques may form the basis for improved characterization of machining processes. Such tools would have great value in process monitoring and quality control.
9. We studied only two of many possible sampling plans. Two other sampling plans deserve to be studied. One is the common industry practice of equal-spaced sampling. (We suspect that this sampling plan is quite poor.) The other sampling plan is actually a class of plans based on fractional factorial experimental design. (See Box *et al.* [5].) (Such plans, which have proven to be highly efficient in many applications, have never been studied for CMM measurement. We suspect that plans based on fractional factorial experimental design will produce even greater efficiencies than the other plans we have studied.)

## 5 ACKNOWLEDGEMENTS

The authors would like to thank the following NIST staff members for their help. Ted Vorburger supplied the roughness reference artifacts. Ralph Veale provided the NAS test part. Alkan Donmez provided the optical flat. Bruce Borchard helped decipher the operation of the CMM. Howard Harary provided many valuable comments on an earlier draft of this manuscript.

## 6 REFERENCES

- [1] Shaw C. Feng and Theodore H. Hopp, "A Review of Current Geometric Tolerancing Theories and Inspection Data Analysis Algorithms", NISTIR 4509, February 1991.
- [2] Filliben, James J., "Dataplot—An Interactive High-Level Language for Graphics, Non-Linear Fitting, Data Analysis, and Mathematics," *Computer Graphics*, 15(3):199-213, 1981.
- [3] Tukey, John W., *Exploratory Data Analysis*, Addison-Wesley, Reading, MA, 1977.
- [4] Chambers, John M., Cleveland, William S., Kleiner, Beat, and Tukey, Paul A., *Graphical Methods for Data Analysis*, Duxbury Press, Boston, MA, 1983.
- [5] Box, Hunter and Hunter, *Statistics for Experimenters*, Wiley, NY, 1978.

# The Control Architecture for Quality in Automation

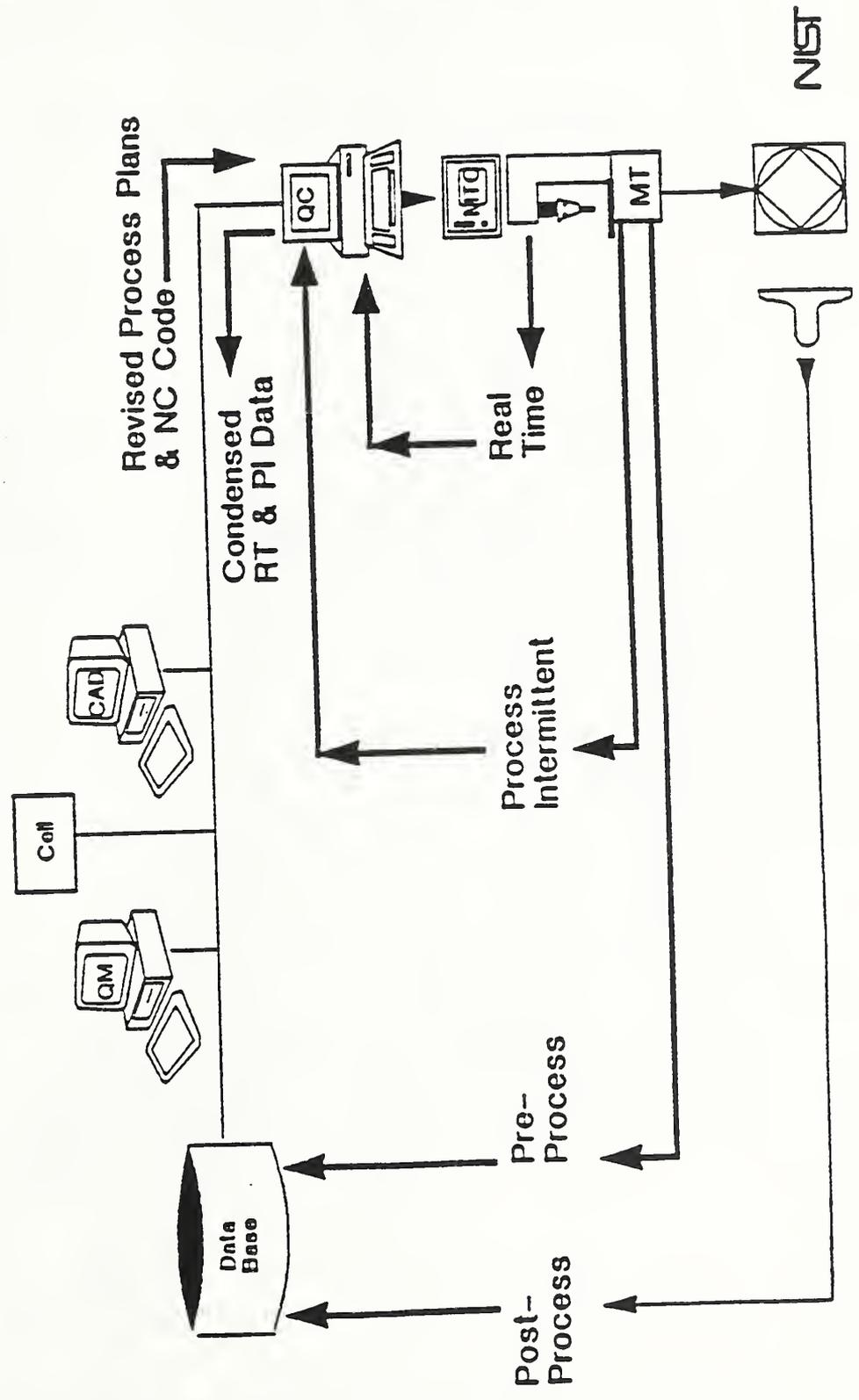
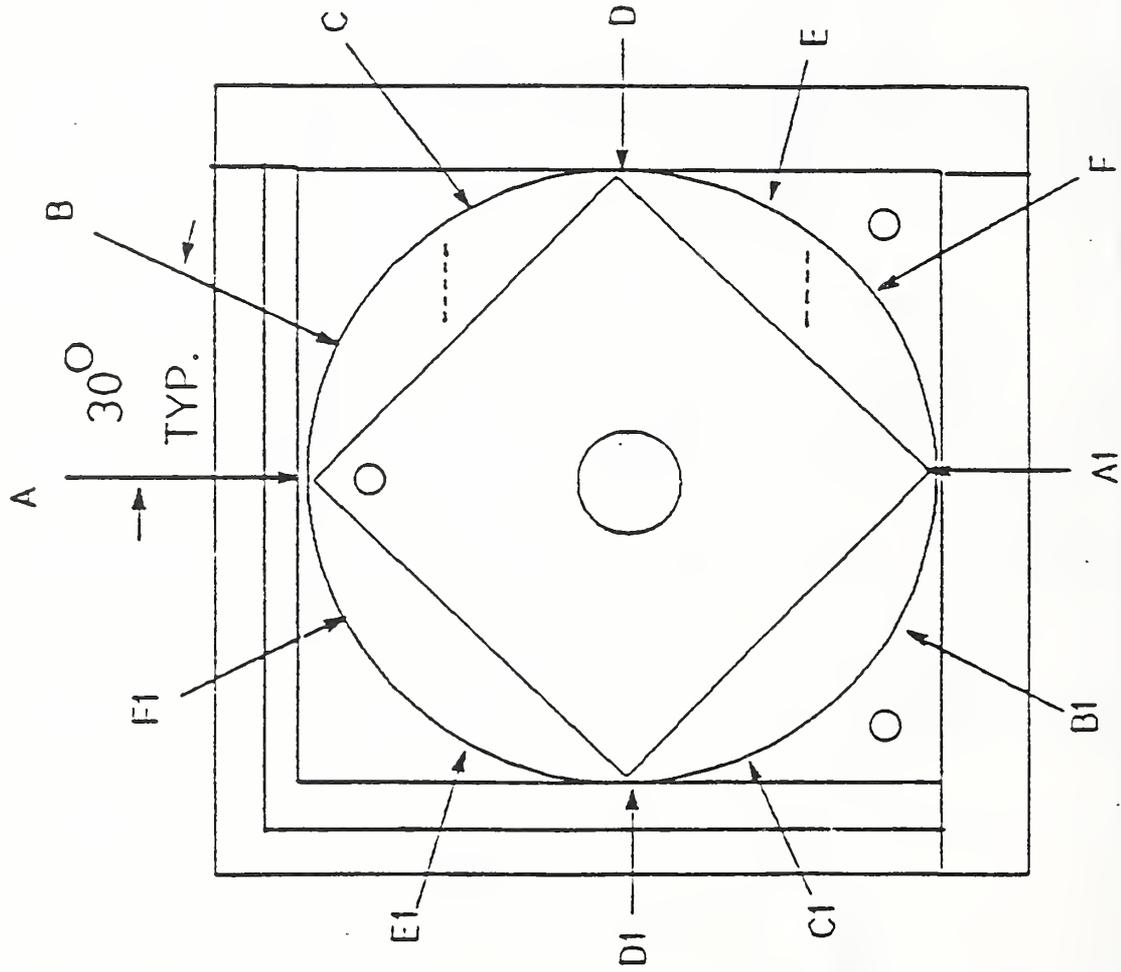


Figure 1. Quality in Automation Control Architecture

# NAS 979 CUTTING TEST PART NUMBER HWS-1



POINT	RADIUS
A	5.9997
B	5.9992
C	5.9990
D	5.9980
E	5.9972
F	5.9973
A1	5.9976
B1	5.9986
C1	5.9990
D1	5.9992
E1	5.9994
F1	5.9996

Figure 2. NAS 979 Artifact

# GENERAL ELECTRIC STANDARD ROUGHNESS SPECIMENS ©

Cat. No. 8651831G1

Symbol	General Electric Symbol	Roughness Micro-Inches		Typical Methods of Producing (There are additional methods not listed here.)
		Average	Average Peak to Valley	
A	$r^1$	4	15	Ground, lapped, honed, micro-honed, superfinished, etc.
B	$r^2$	8	28	Ground, lapped, honed, micro-honed, superfinished, molded, drawn, etc.
C	$r^3$	10	16	Ground, lapped, honed, reamed, scraped, superfinished, bored, burnished, planished, milled, molded, precision-cast, cold-pressed, drawn, extruded, etc.
D	$r^{3\frac{1}{2}}$	32	118	Ground, broached, reamed, scraped, turned, bored, burnished, milled, planished, molded, precision-cast, cold-pressed, drawn, extruded, etc.
E	$r^4$	63	220	Ground, broached, rolled, shaped, turned, bored, milled, abrasive cut-off, reamed, molded, precision-cast, extruded, drawn, etc.
F	$r^5$	125	455	Ground, broached, rolled, reamed, shaped, turned, bored, milled, drilled, spot-faced, counterboard, filed, punched, abrasive cut-off, molded, precision-cast, extruded, forged, etc.
G	$r^6$	250	875	Ground, rolled, shaped, turned, bored, milled, drilled, spot-faced, counterbored, filed, punched, sand-cast, forged, precision-cast, etc.
H	$r^{6\frac{1}{2}}$	500	1750	Ground, gas-cut, chipped, shaped, turned, bored, milled, sawed, filed, punched, sand-cast, forged, etc.
J	$r^7$	1000	3500	Ground, gas-cut, chipped, shaped, turned, bored, milled, sawed, sheared, nibbled, welded, punched, sand-cast, forged, etc.
K	$r^8$	2000	7000	Gas-cut, chipped, shaped, turned, milled, sawed, shared, nibbled, welded, parting cut, sand-case, etc.

General Electric Standard Roughness Specimens are surfaces of graded roughness for reference and comparison purposes in product specification, production and inspection. Comparisons are generally made by sight and feel methods. Regardless of the material or method of processing used on the finished product, each specimen represents the roughest surface acceptable for each symbol. For best results avoid subjecting the surfaces to rough handling or contact with sharp objects. In general "average" values were determined by a Profilometer, and "average peak to valley" values, by a Brush Surface Analyzer or a profile recorder, except as indicated in Instructions GEJ-1662.

Figure 3. GE Standard Roughness Specimens

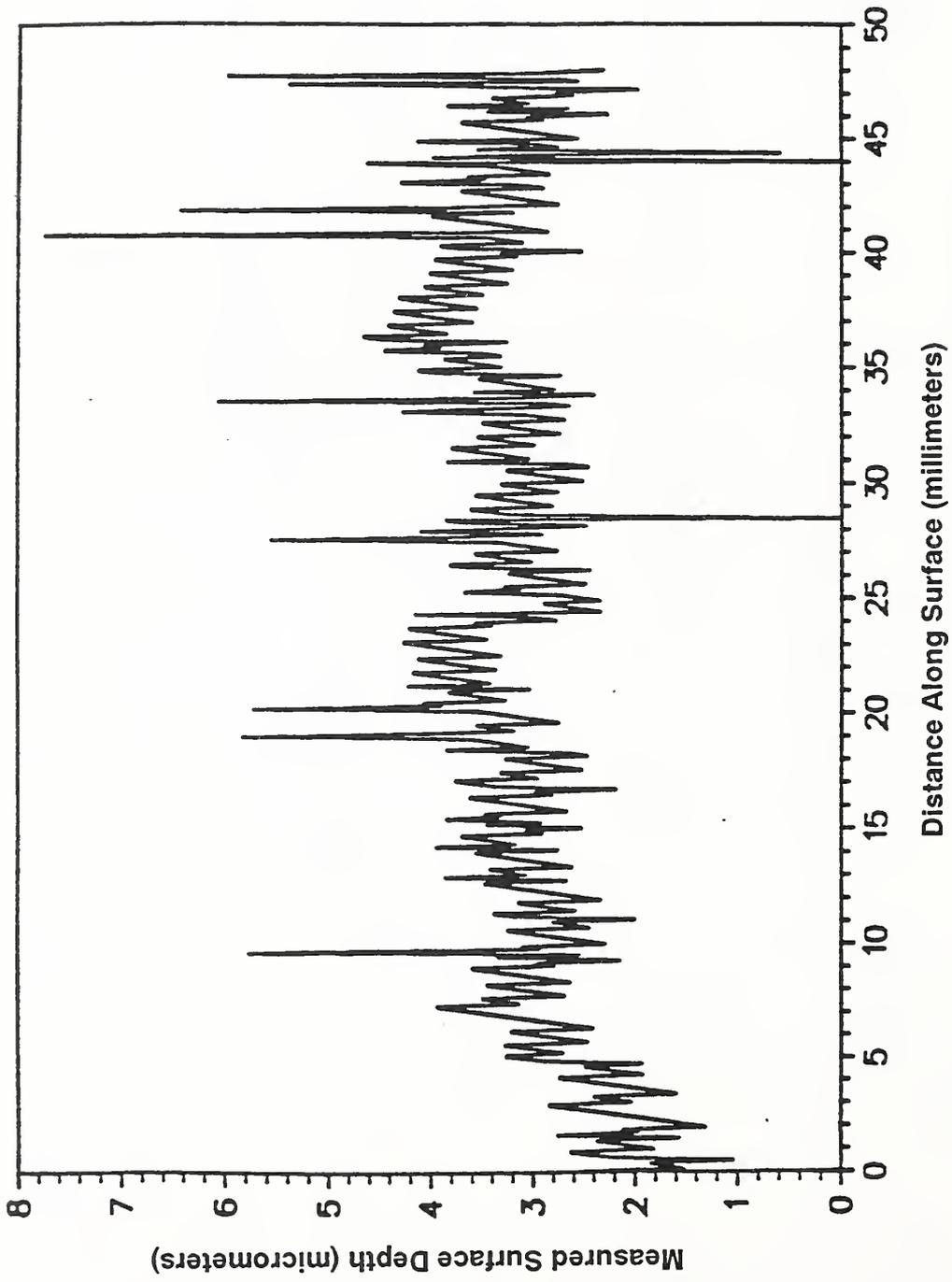
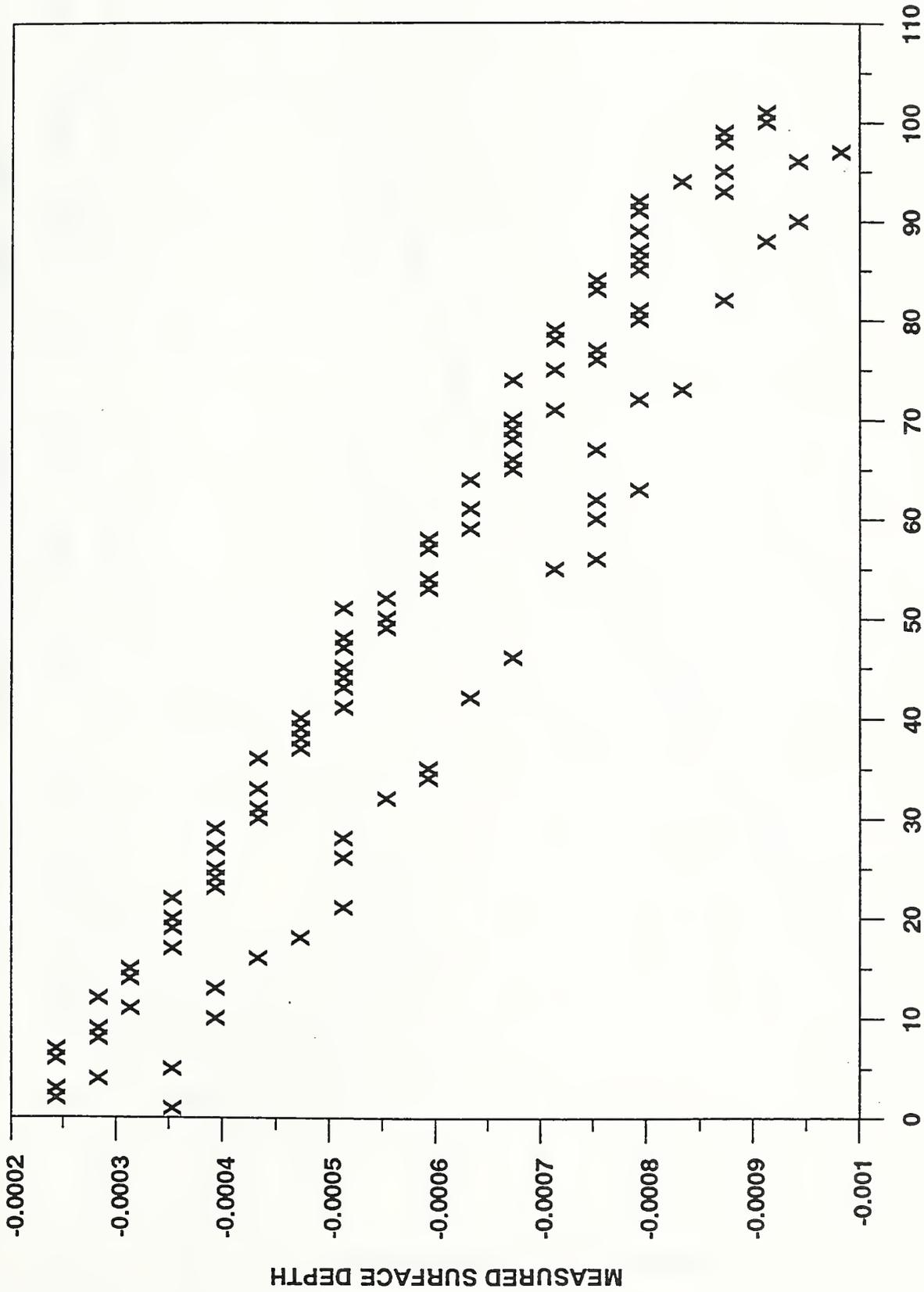
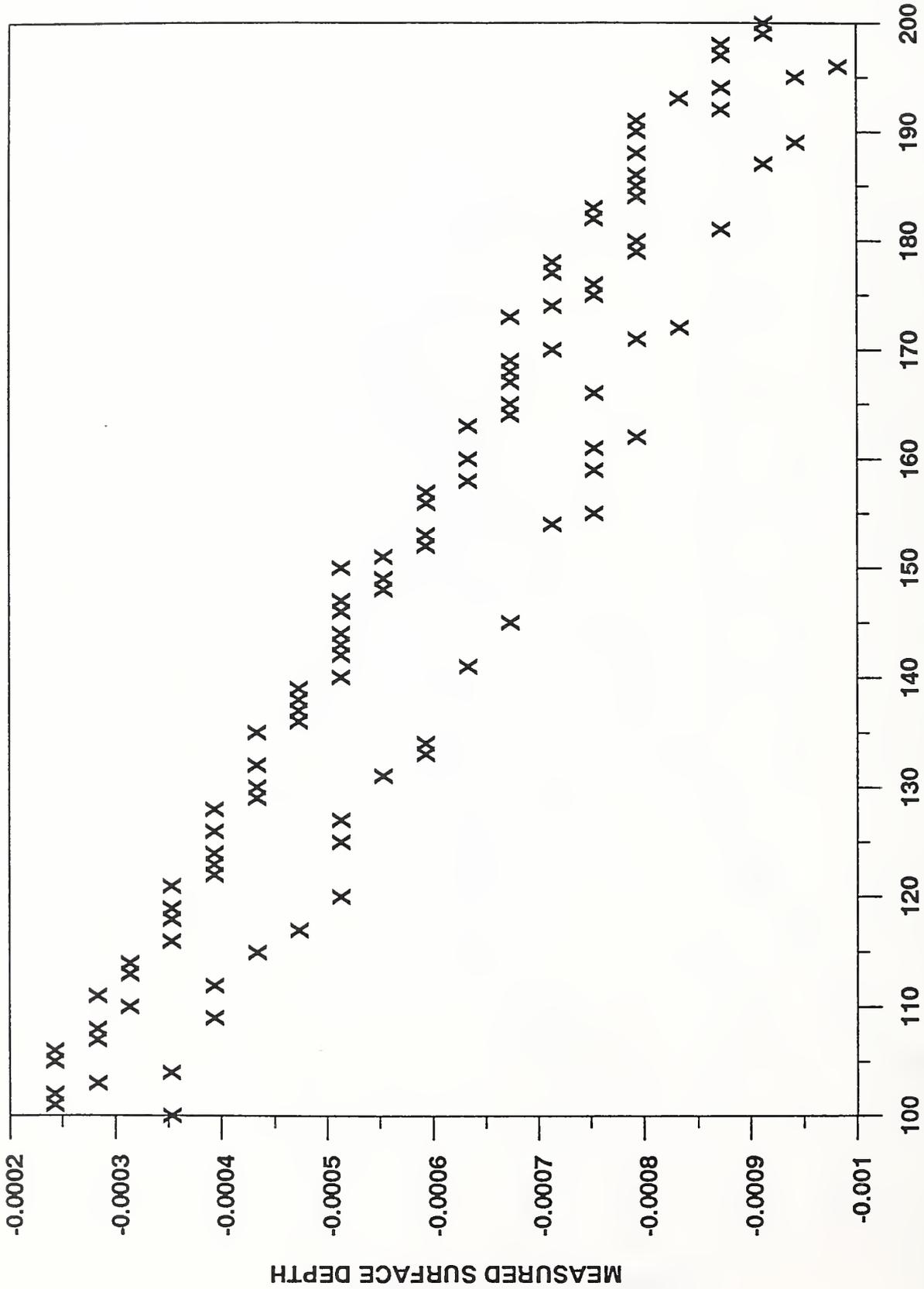


Figure 4. Optical Flat: Surface profile (400 points)



CODED DISTANCE ALONG SURFACE (CODED TOTAL DISTANCE = 400)

Figure 5. Specimen GE-A: First 50 points (out of the 400)  
(Evidence of Ram Oscillation Effect)



CODED DISTANCE ALONG SURFACE (CODED TOTAL DISTANCE = 400)

Figure 6. Specimen GE-A: Points 100 to 200 (out of the 400)  
(Evidence of Ram Oscillation Effect)

# SPECIMEN GE-A-4-15 COMPLETELY RANDOMIZED DESIGN

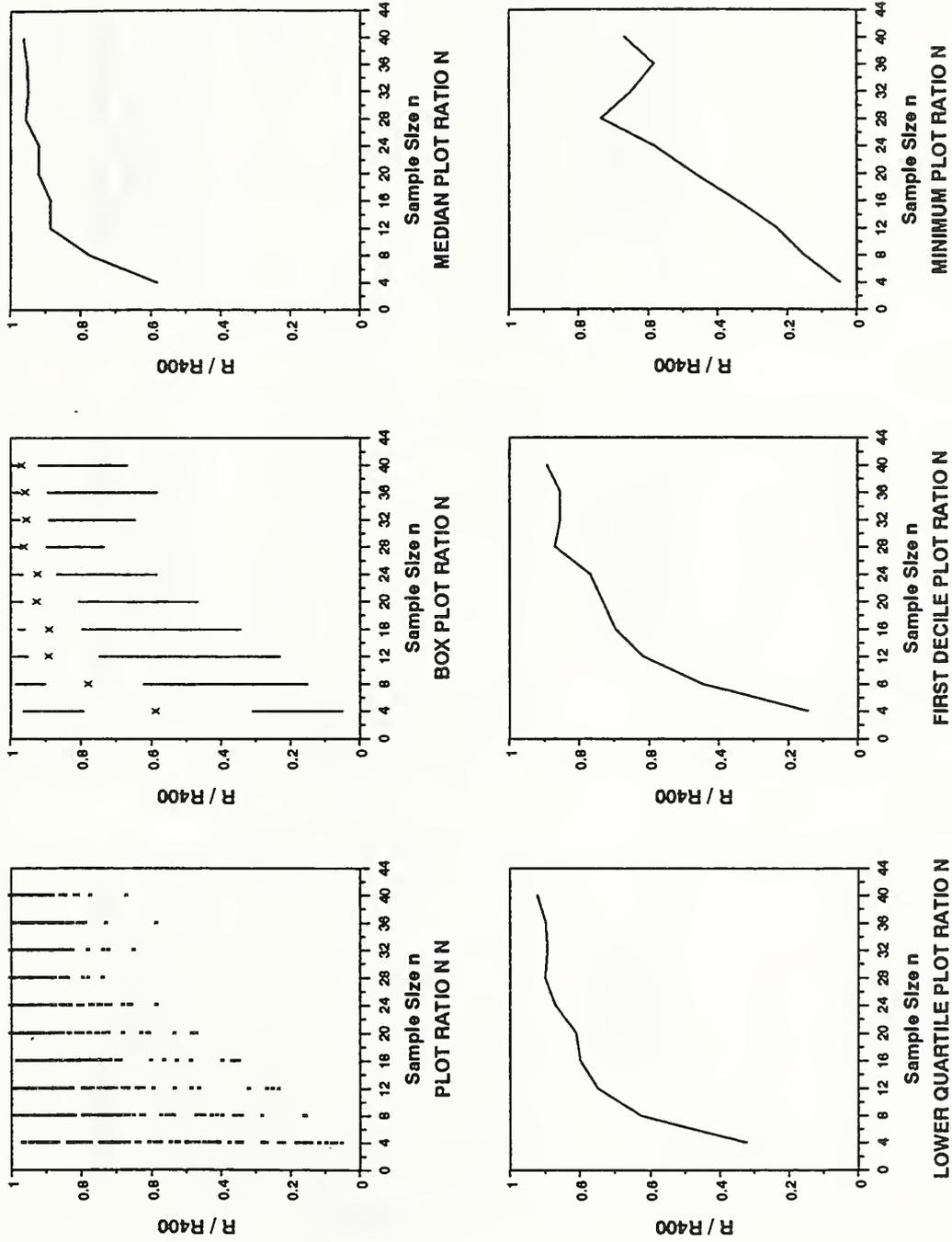


Figure 7-a. GE-A: distribution of  $R/R_{400}$ ; completely randomized design

# SPECIMEN GE-B-8-28 COMPLETELY RANDOMIZED DESIGN

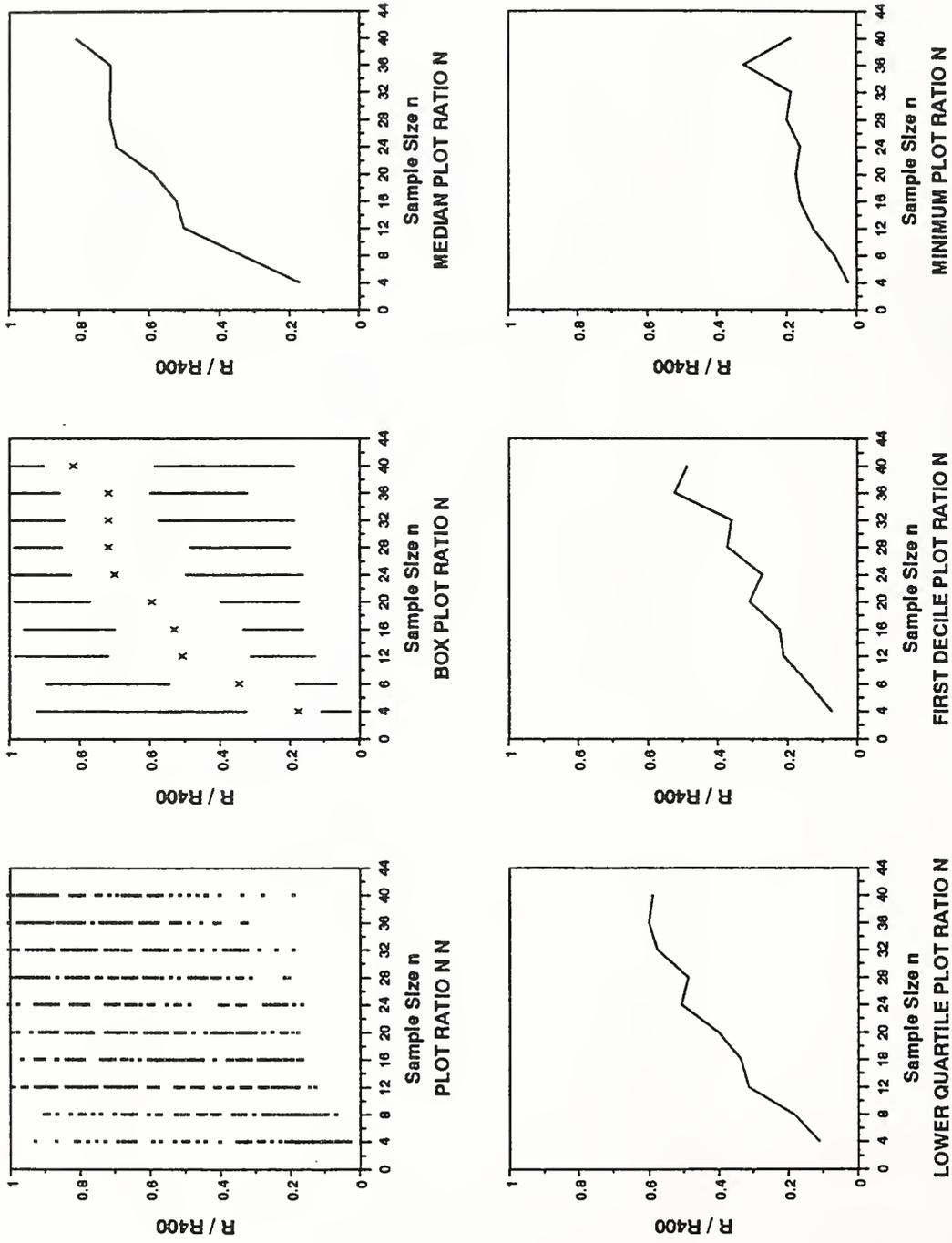


Figure 7-b. GE-B: distribution of  $R/R_{400}$ ; completely randomized design

# SPECIMEN GE-C-16-56 COMPLETELY RANDOMIZED DESIGN

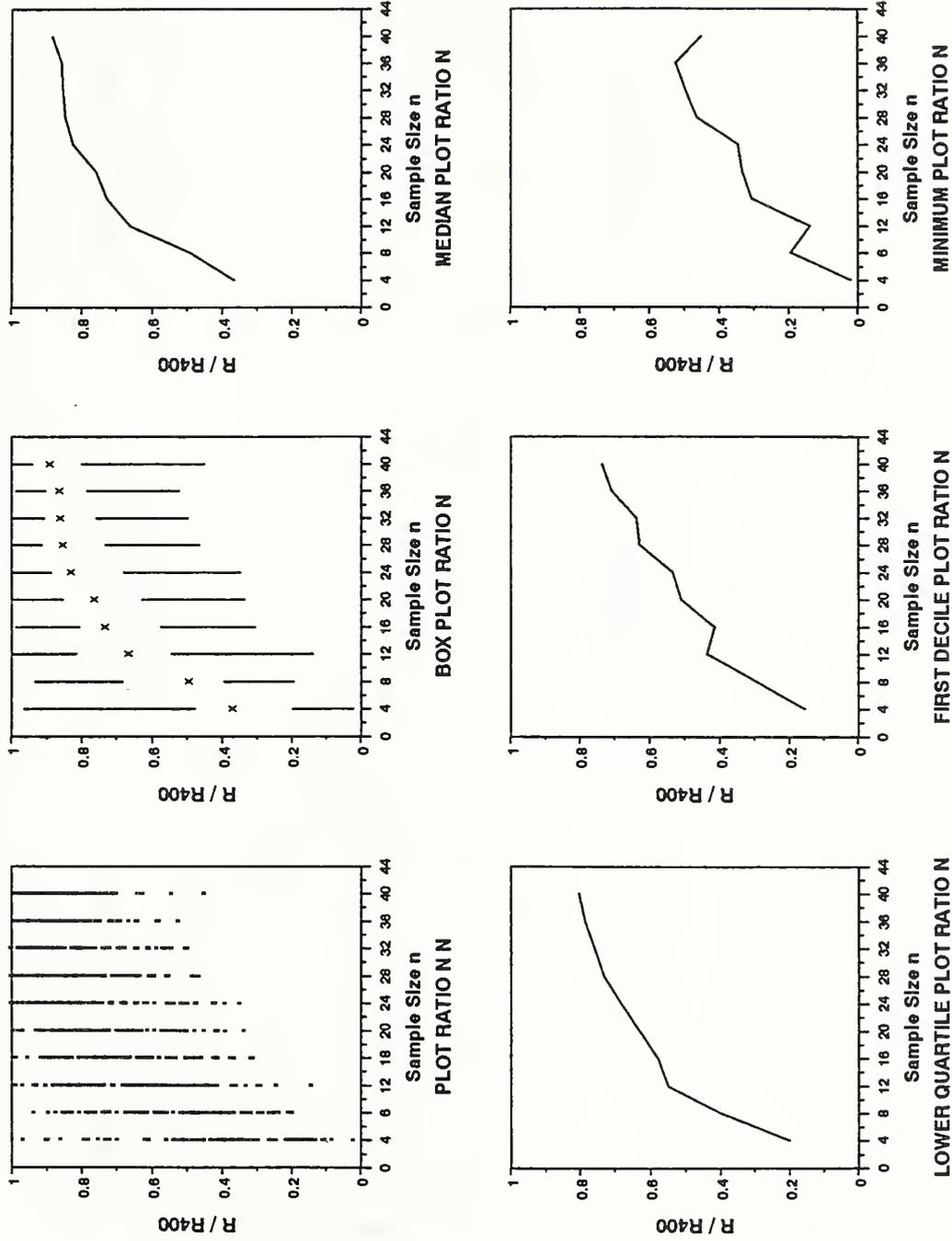


Figure 7-c. GE-C: distribution of  $R/R_{400}$ ; completely randomized design

# SPECIMEN GE-D-32-118 COMPLETELY RANDOMIZED DESIGN

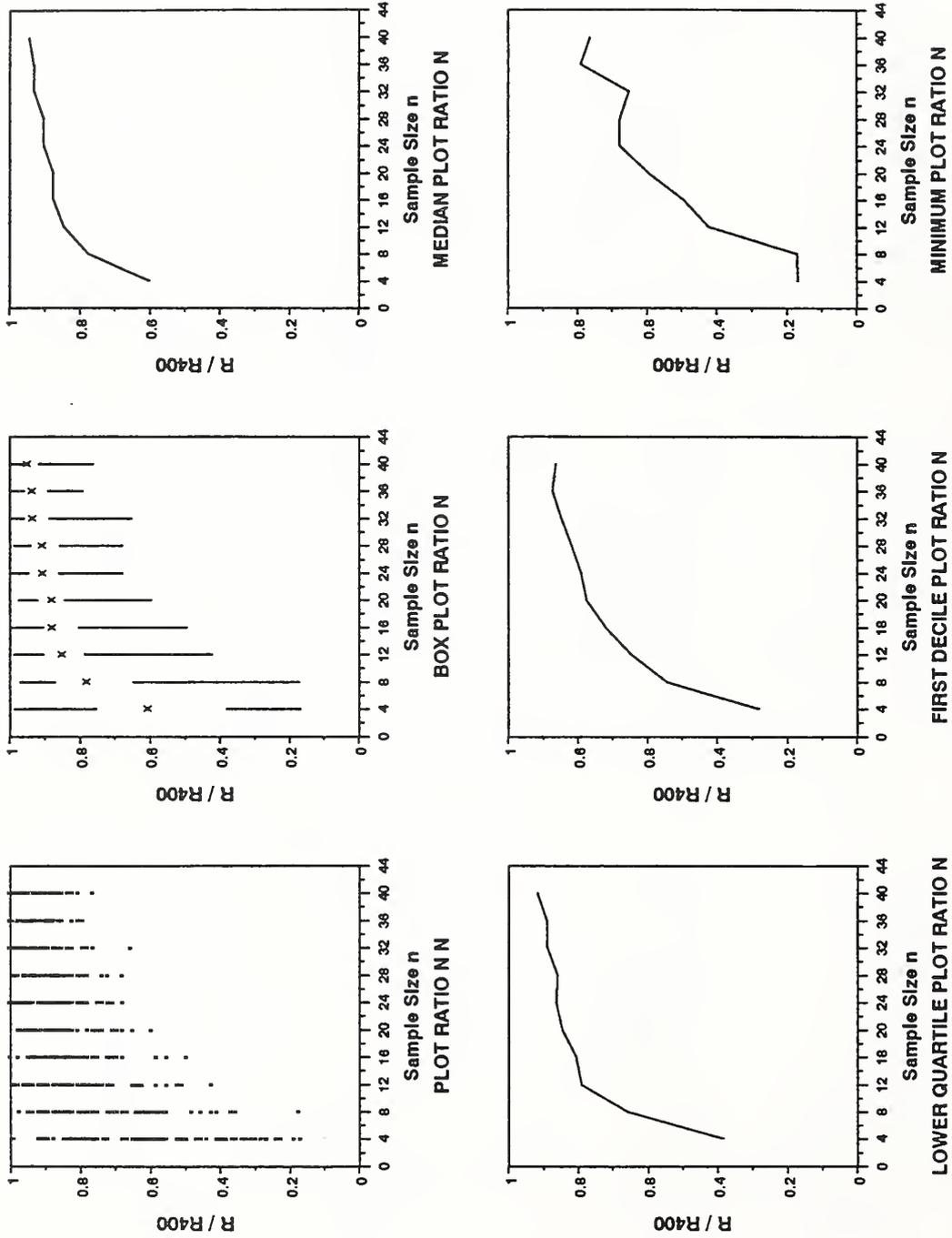


Figure 7-d. GE-D: distribution of  $R/R_{400}$ ; completely randomized design

# SPECIMEN GE-E-63-220 COMPLETELY RANDOMIZED DESIGN

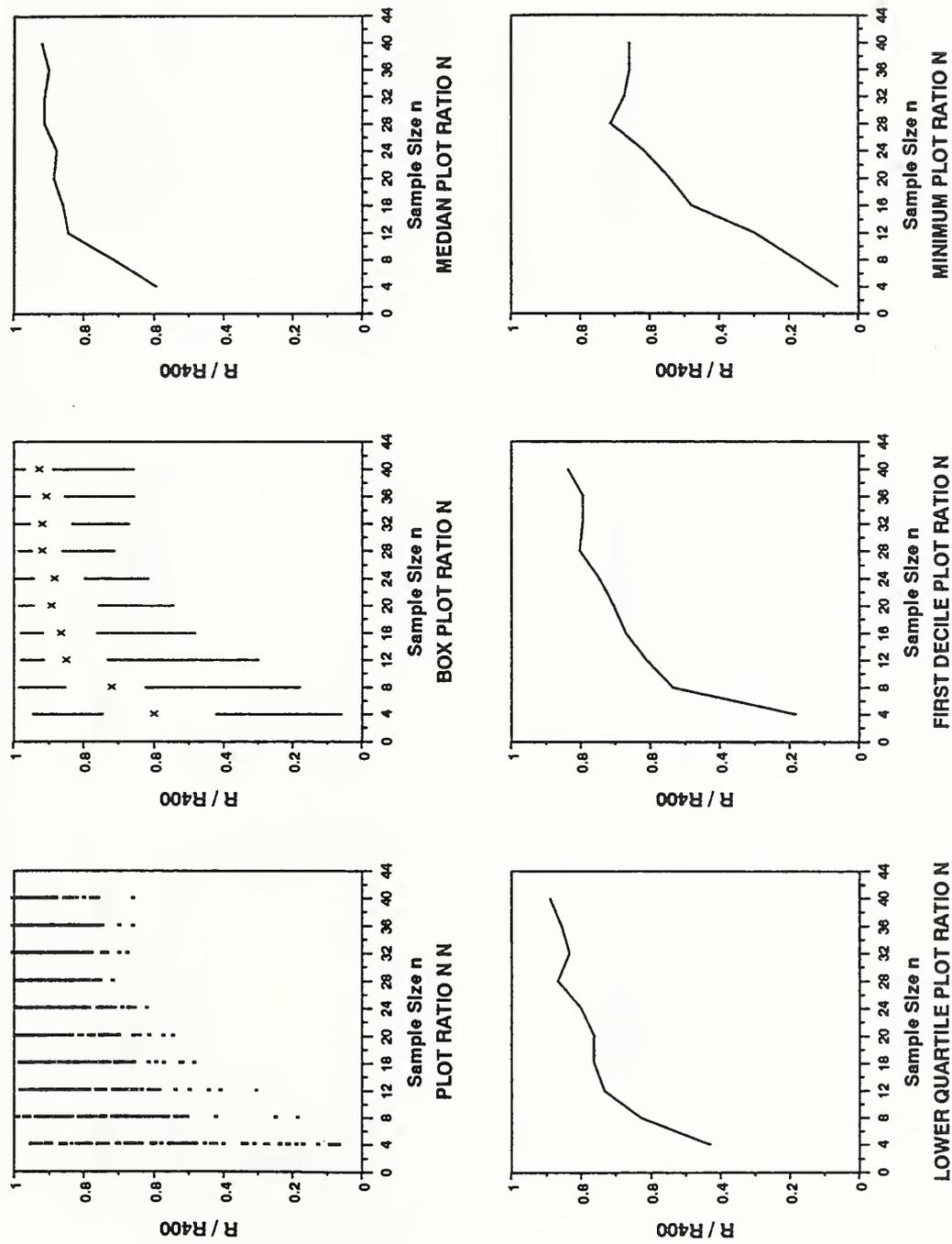


Figure 7-e. GE-E: distribution of  $R/R_{400}$ ; completely randomized design

**SPECIMEN GE-F-125-455 COMPLETELY RANDOMIZED DESIGN**

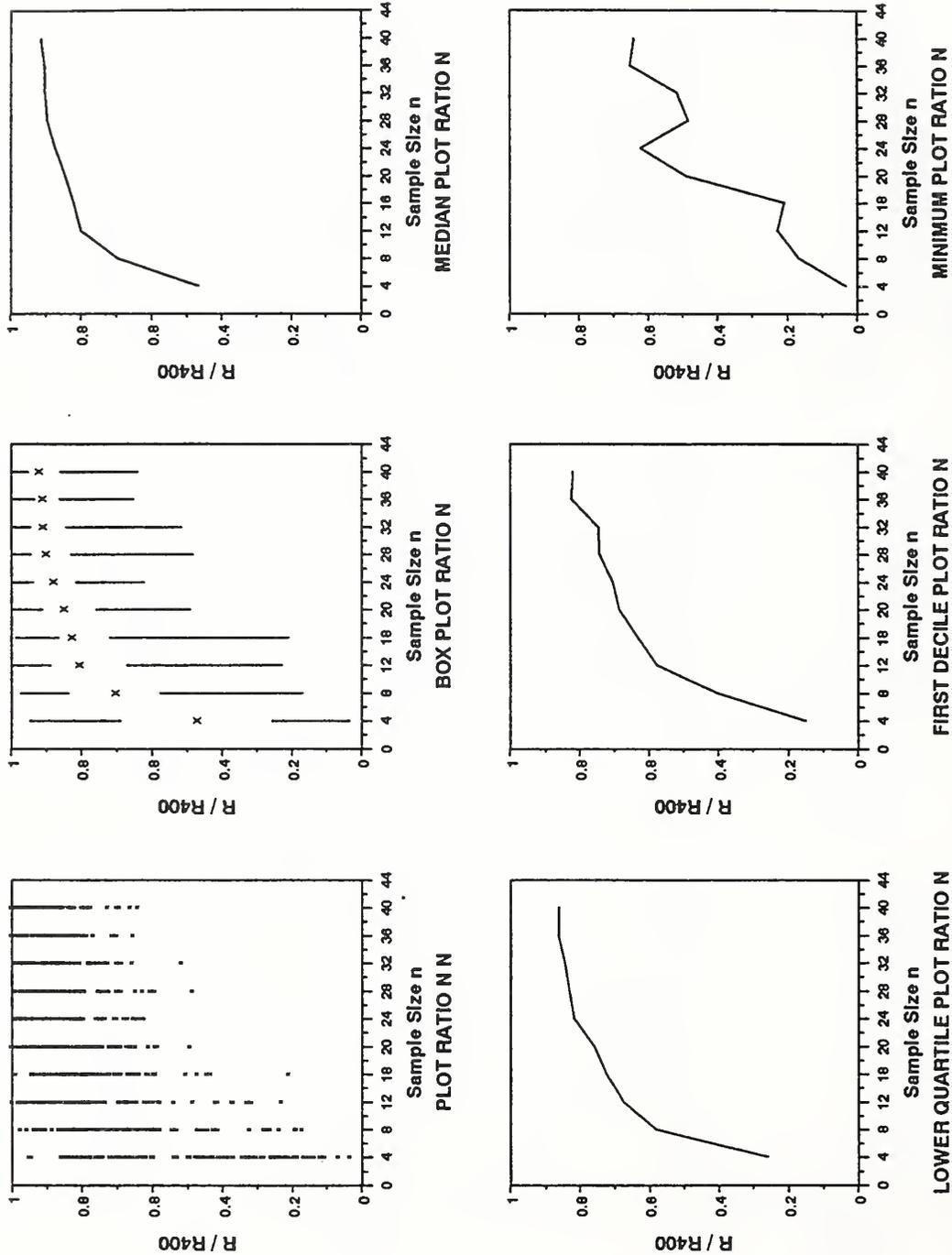


Figure 7-f. GE-F: distribution of  $R/R_{400}$ ; completely randomized design

# SPECIMEN GE-G-250-875 COMPLETELY RANDOMIZED DESIGN

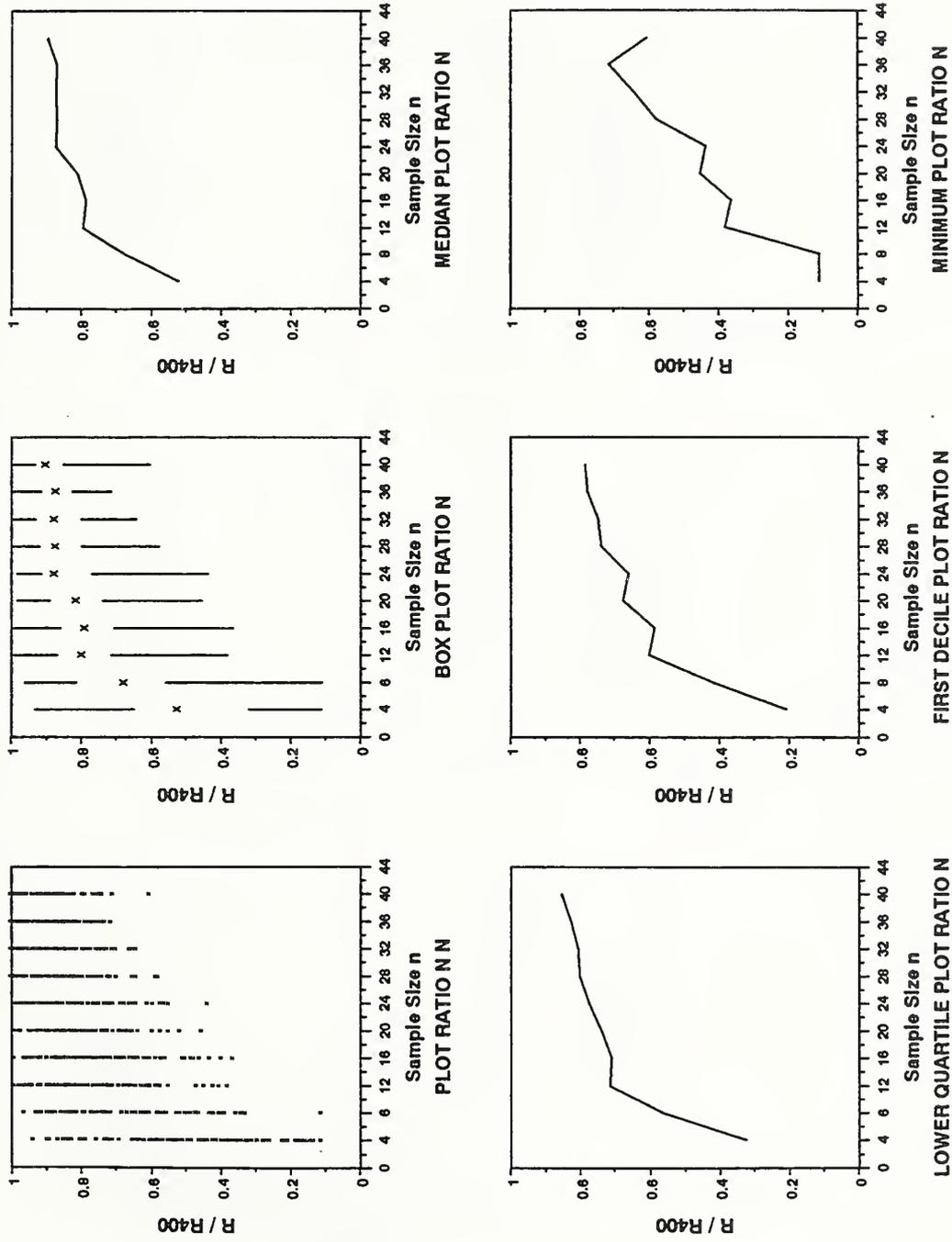


Figure 7-g. GE-G: distribution of  $R/R_{400}$ ; completely randomized design

# SPECIMEN GE-H1-500-1750 COMPLETELY RANDOMIZED DESIGN

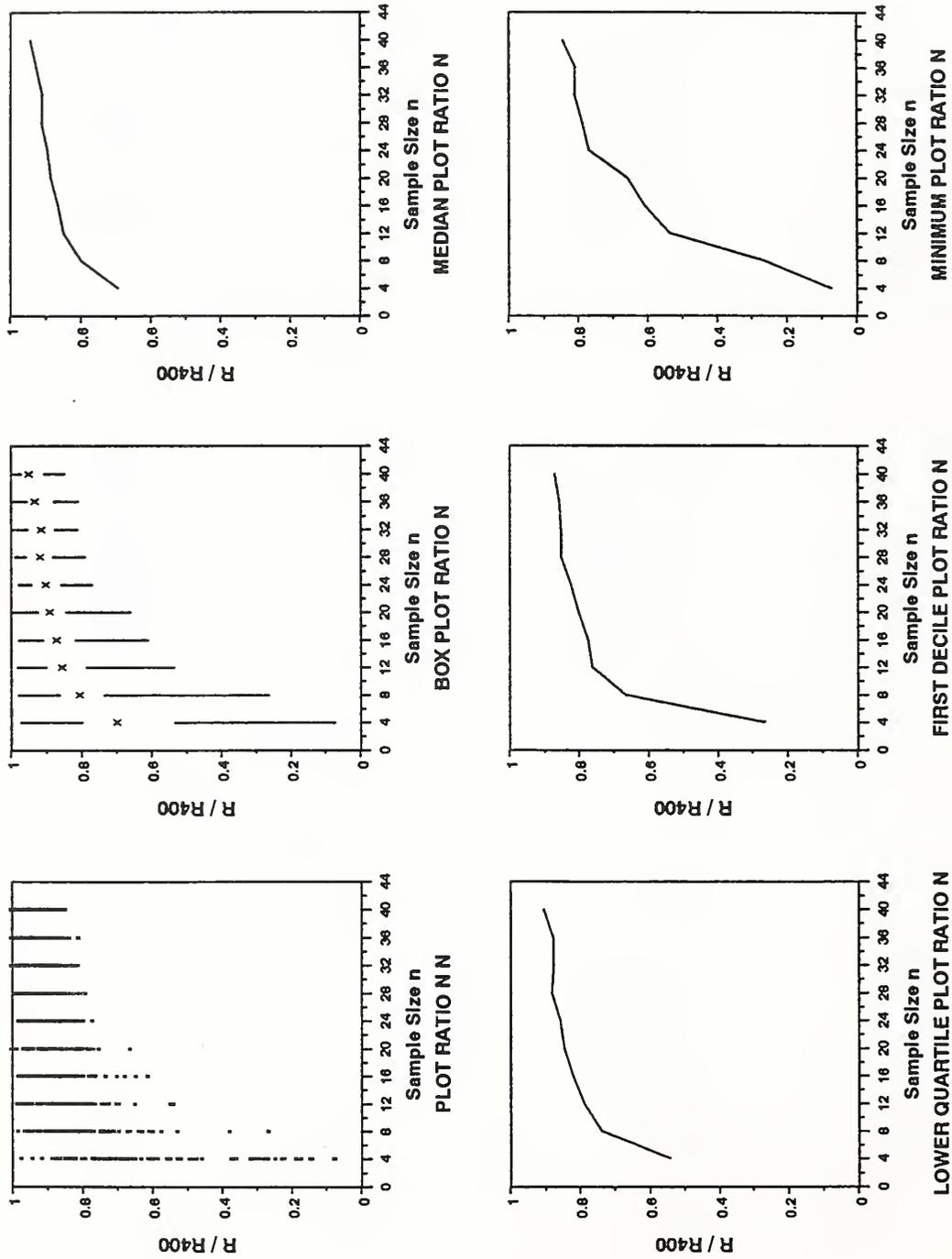


Figure 7-h. GE-H1: distribution of  $R/R_{400}$ ; completely randomized design

# SPECIMEN GE-H2-500-1750 COMPLETELY RANDOMIZED DESIGN

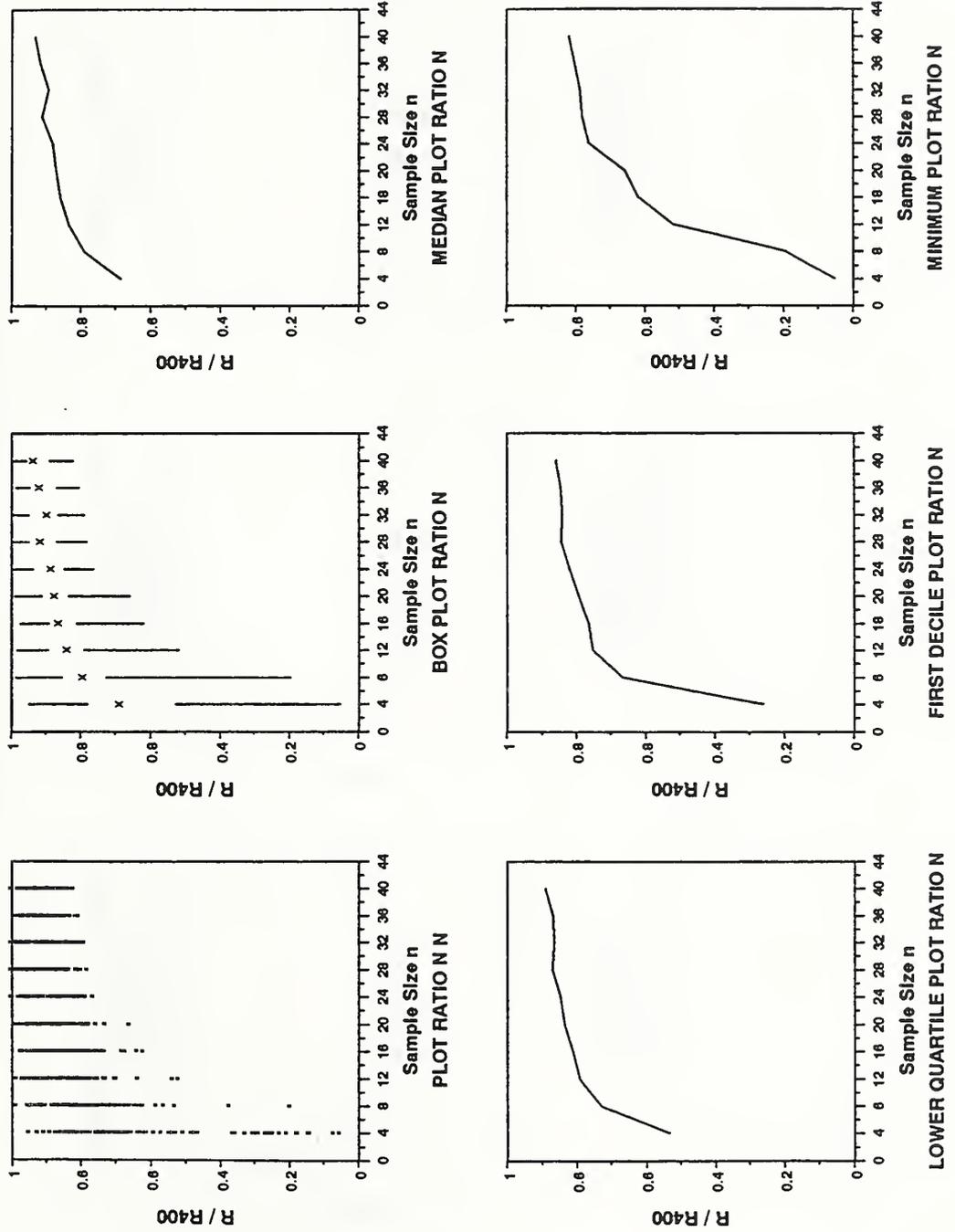


Figure 7-i. GE-H2: distribution of  $R/R_{400}$ ; completely randomized design

# SPECIMEN GE-OPT FLAT COMPLETELY RANDOMIZED DESIGN

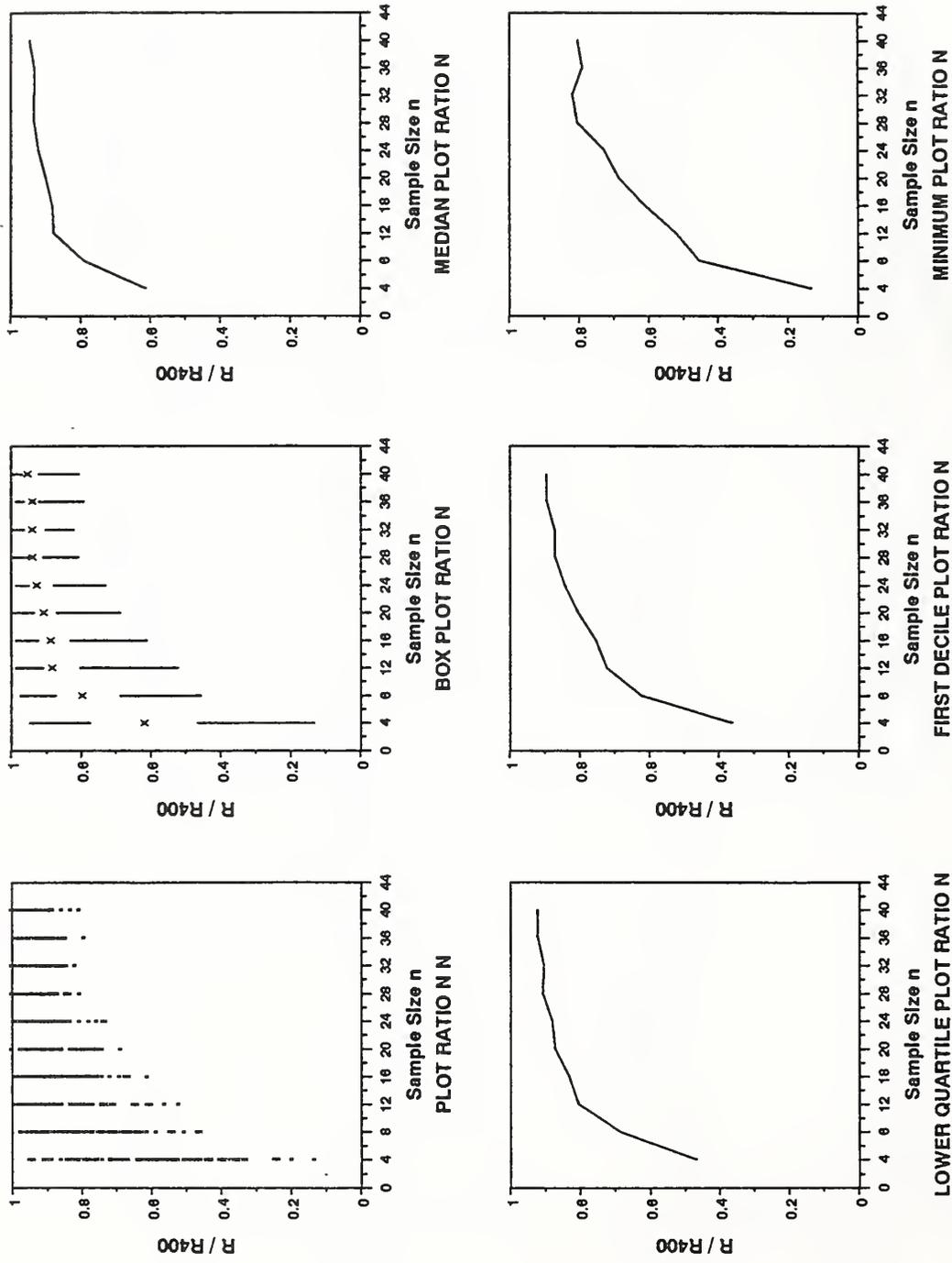


Figure 7-j. Optical Flat: distribution of R/R400; completely randomized design

# SPECIMEN GE-NAS-1 COMPLETELY RANDOMIZED DESIGN

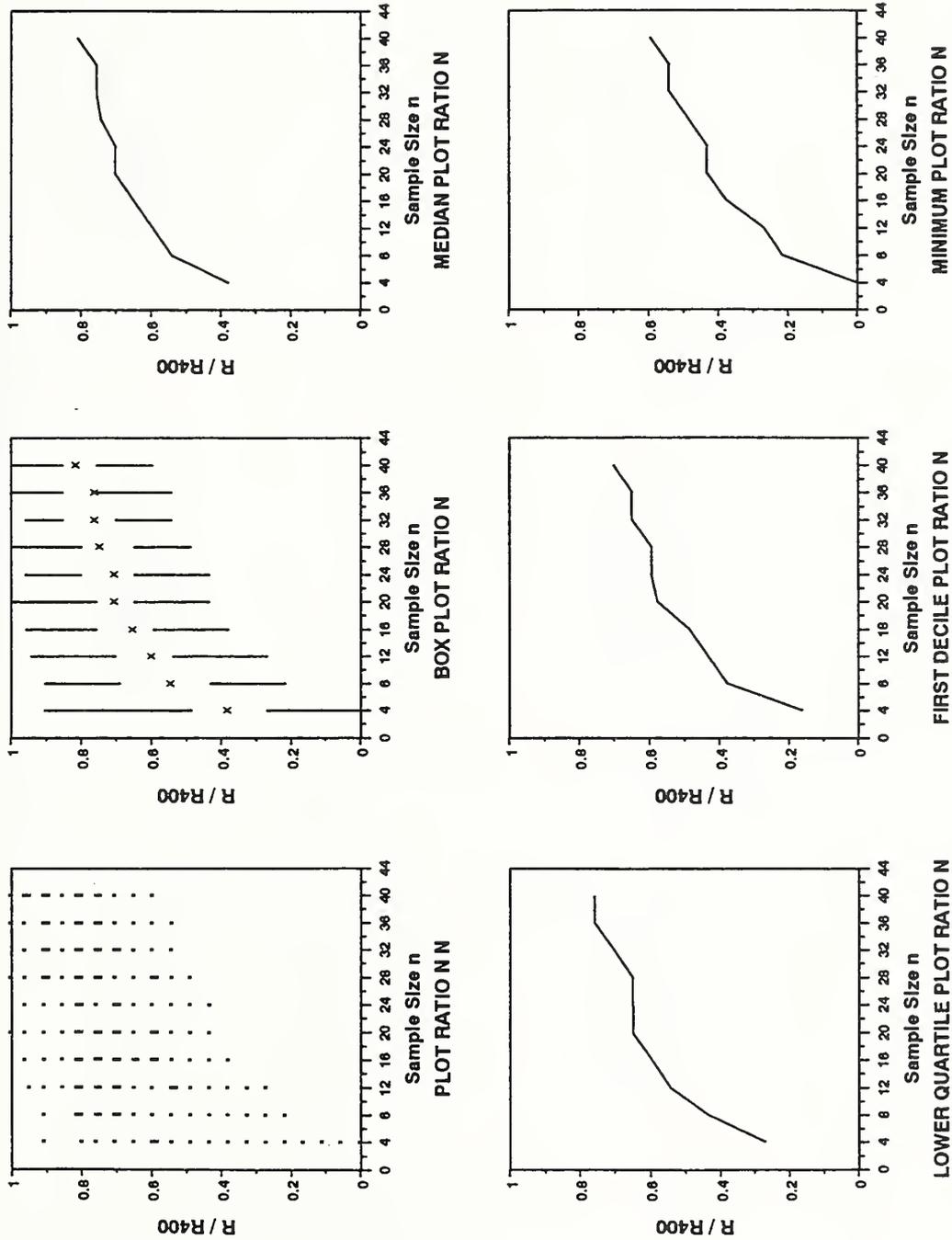


Figure 7-k. NAS-1: distribution of  $R/R400$ ; completely randomized design

# SPECIMEN GE-NAS-2 COMPLETELY RANDOMIZED DESIGN

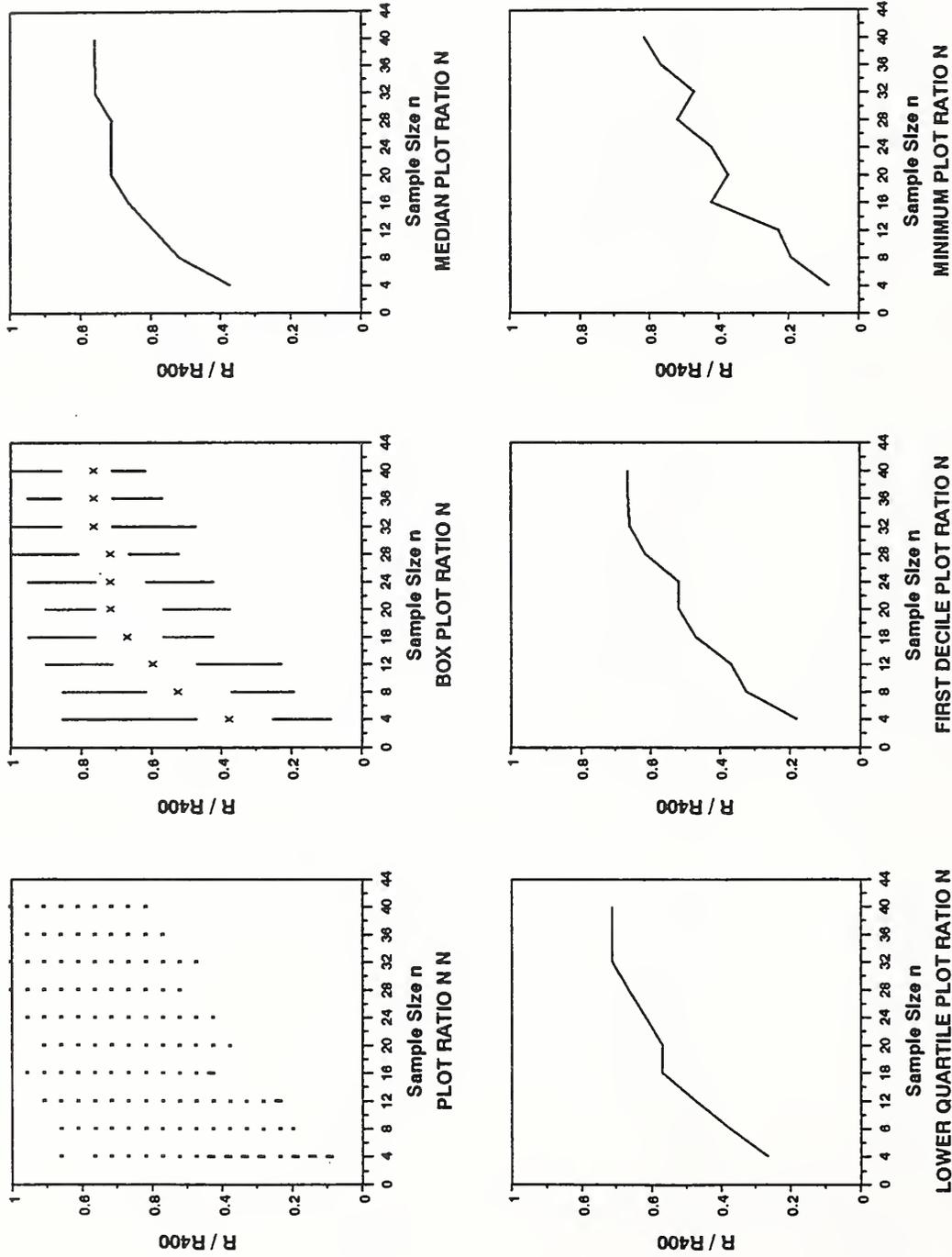


Figure 7-1. NAS-2: distribution of  $R/R_{400}$ ; completely randomized design

# COMPLETELY RANDOMIZED DESIGN--BOX PLOTS

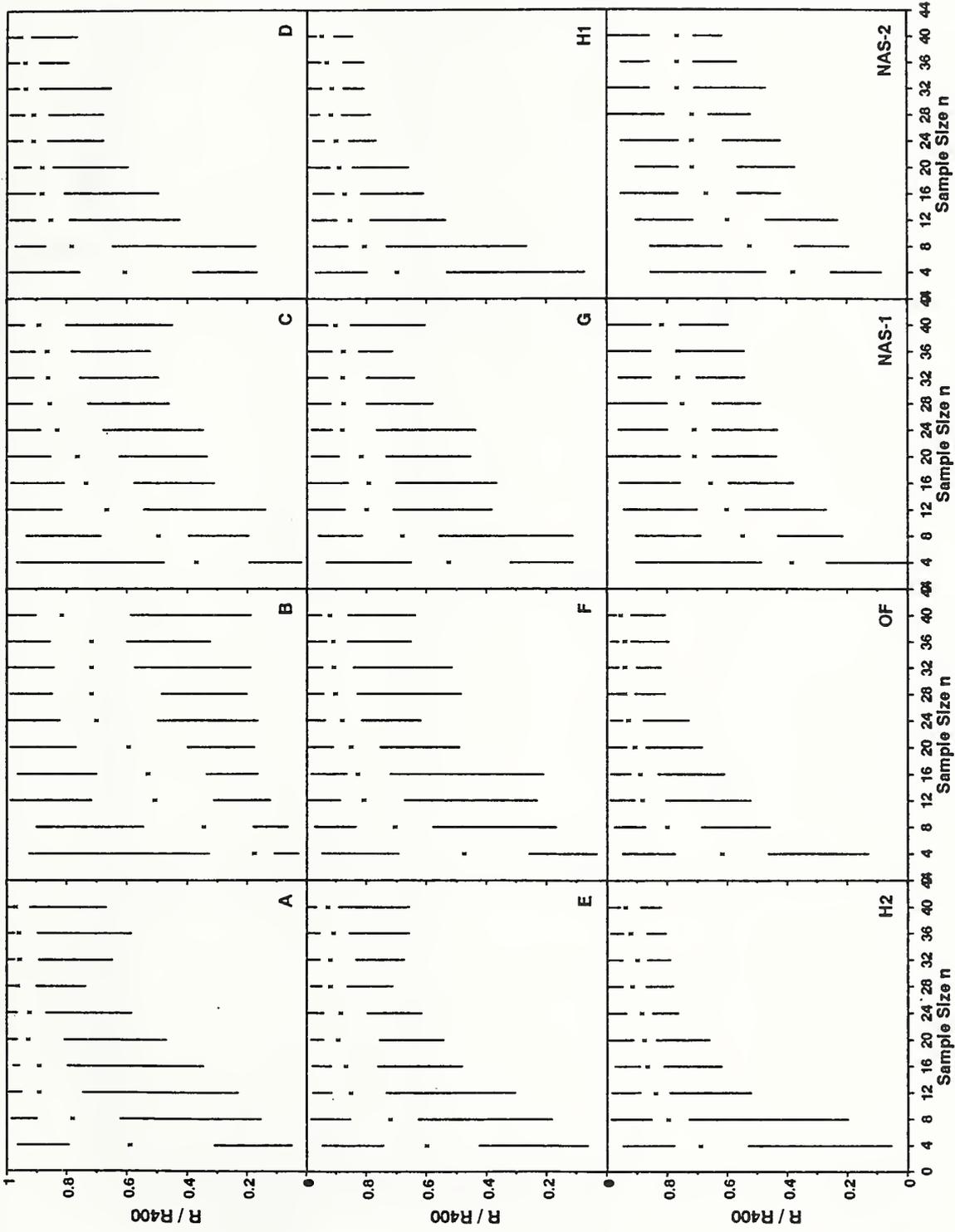


Figure 8-a. Box plots of  $R/R400$  as a function of sample size; completely randomized design

# COMPLETELY RANDOMIZED DESIGN--MEDIAN PLOTS

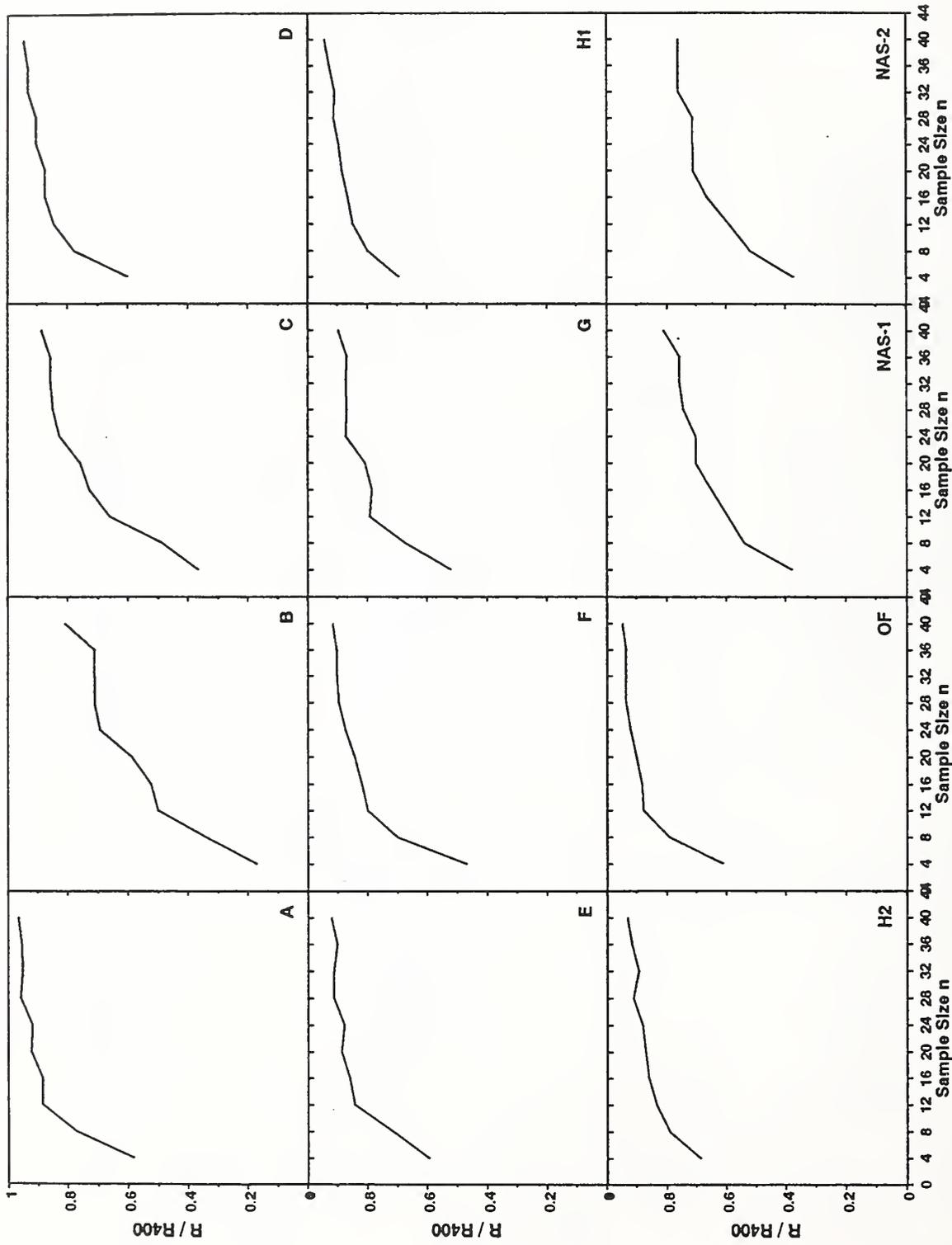


Figure 8-b. Median of  $R/R400$  as a function of sample size;  
completely randomized design

# COMPLETELY RANDOMIZED DESIGN--FIRST DECILE PLOTS

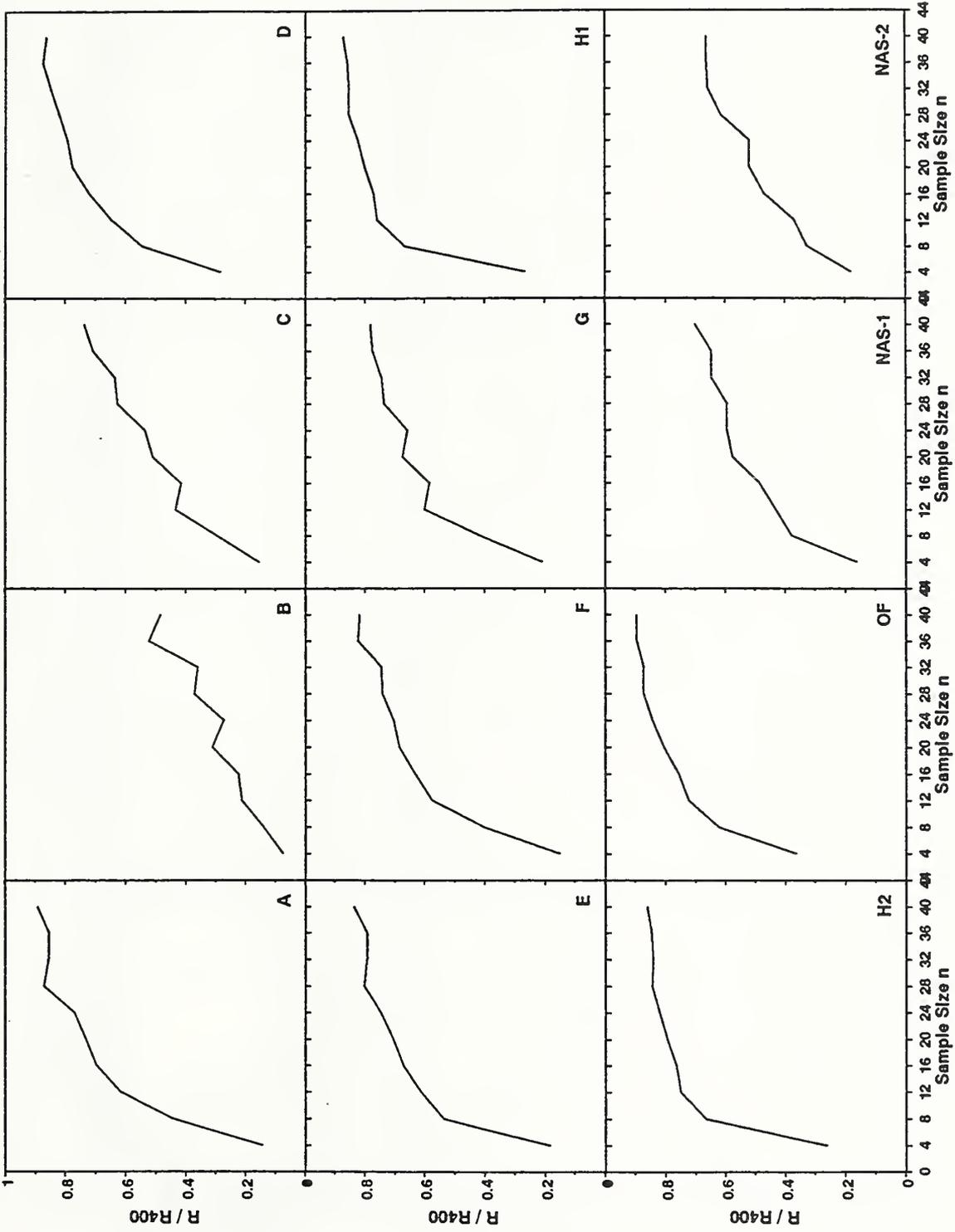


Figure 8-c. First decile of  $R/R400$  as a function of sample size; completely randomized design

# COMPLETELY RANDOMIZED DESIGN--MINIMUM PLOTS

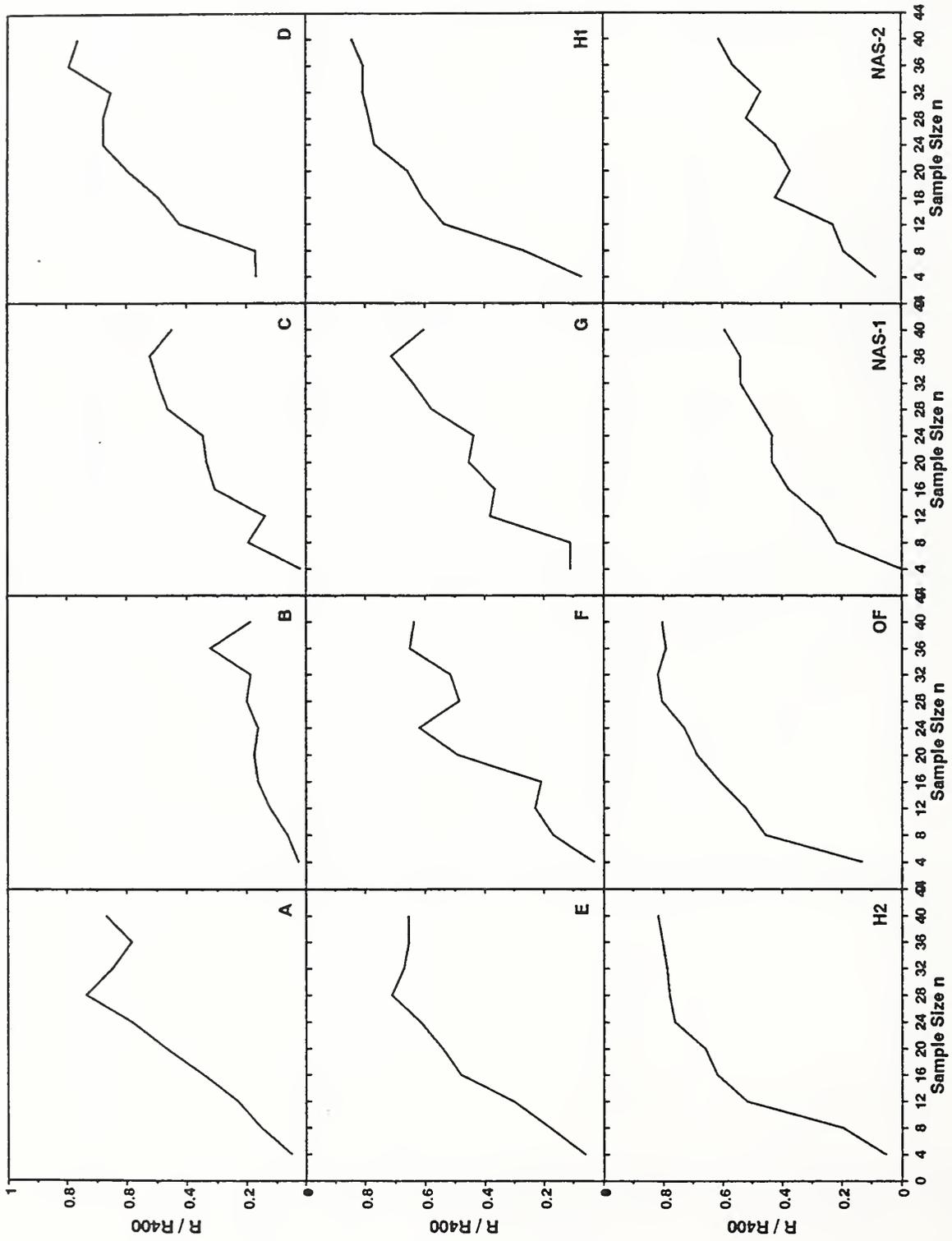


Figure 8-d. Minimum of  $R/R400$  as a function of sample size; completely randomized design

# RANDOMIZED BLOCK DESIGN (4 BLOCKS)--BOX PLOTS

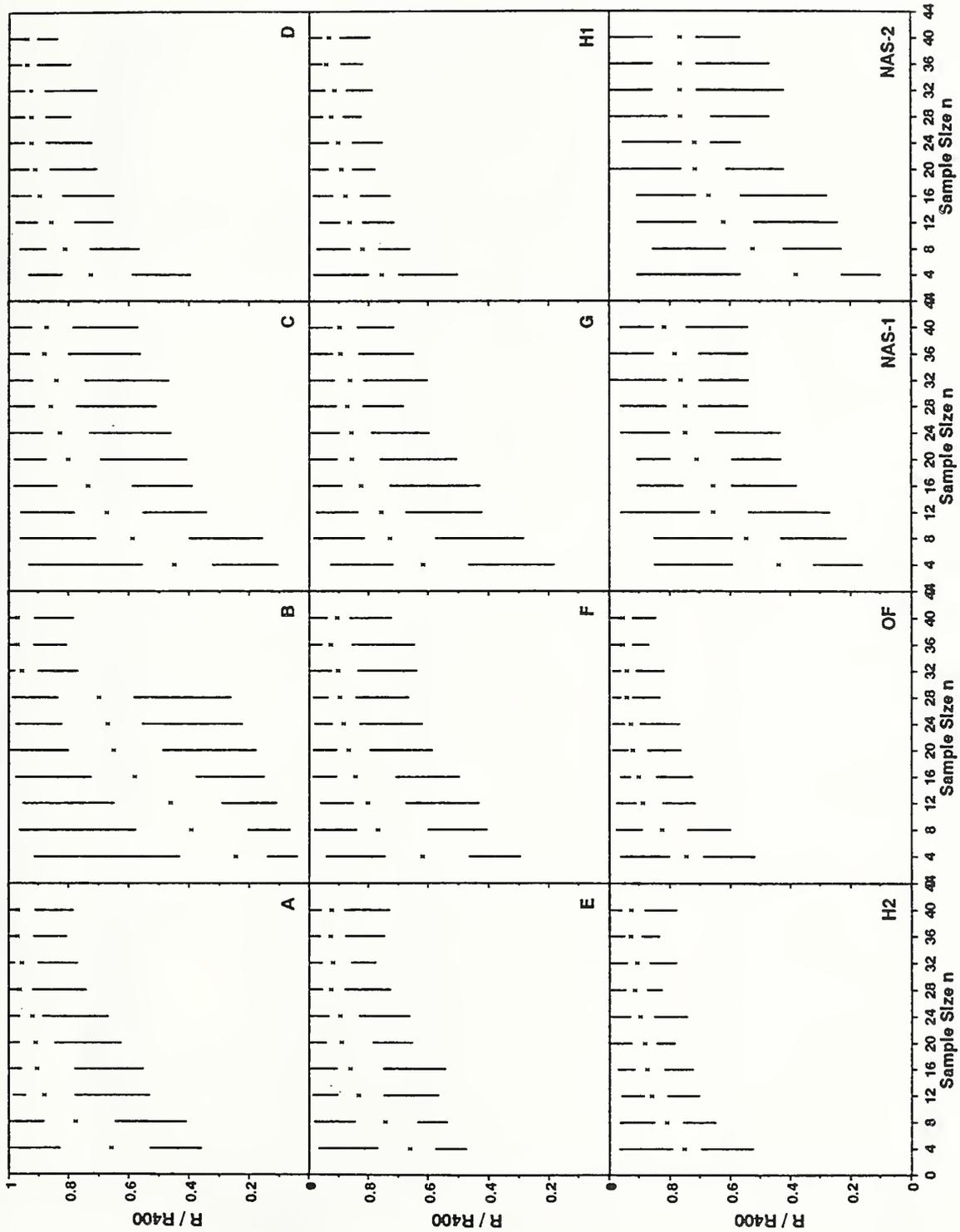


Figure 9-a. Box plots of  $R/R400$  as a function of sample size; randomized block design (4 blocks)

# RANDOMIZED BLOCK DESIGN (4 BLOCKS)--MEDIAN PLOTS

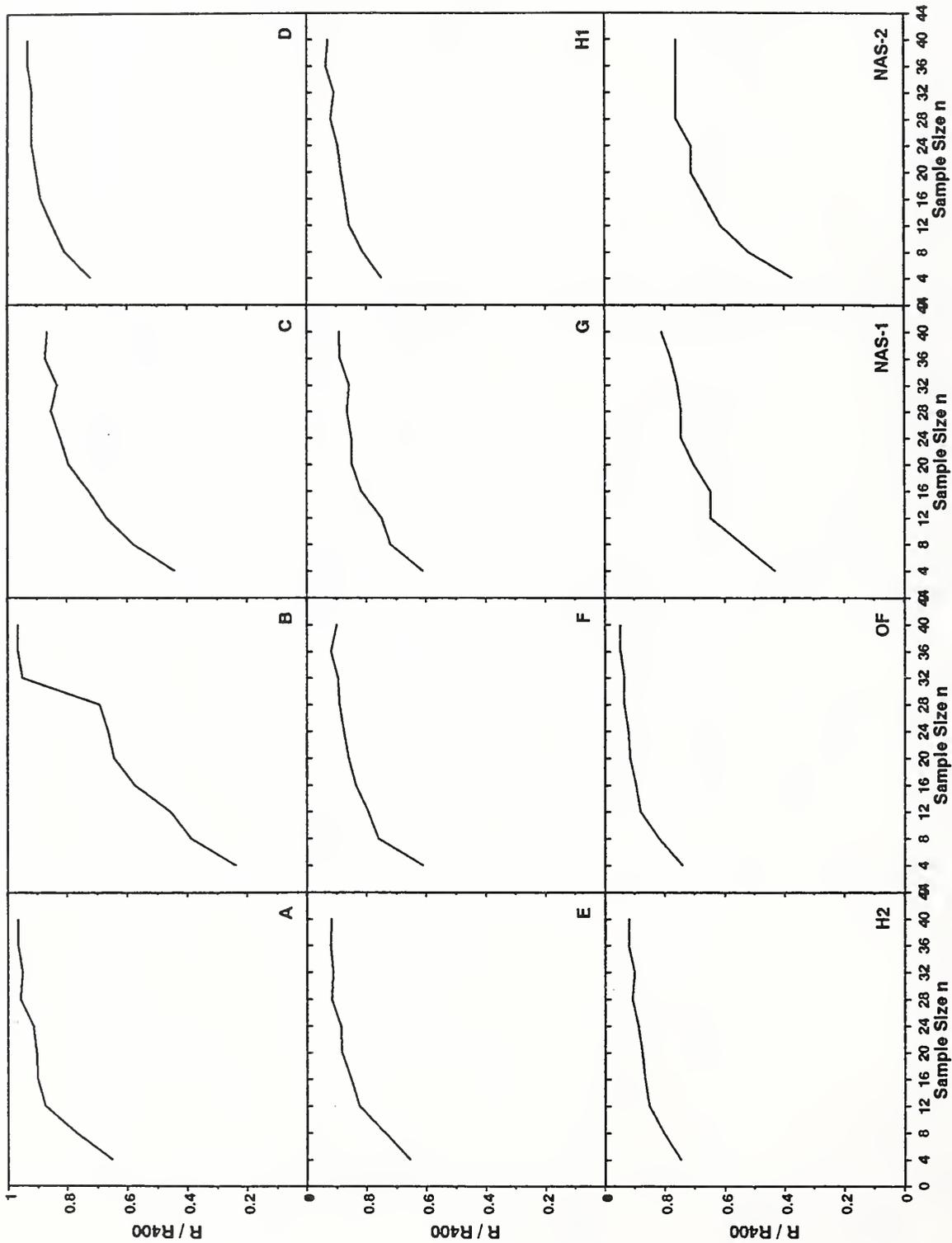


Figure 9-b. Median of R/R400 as a function of sample size; randomized block design (4 blocks)

# RANDOMIZED BLOCK DESIGN (4 BLOCKS)--FIRST DECILE PLOTS

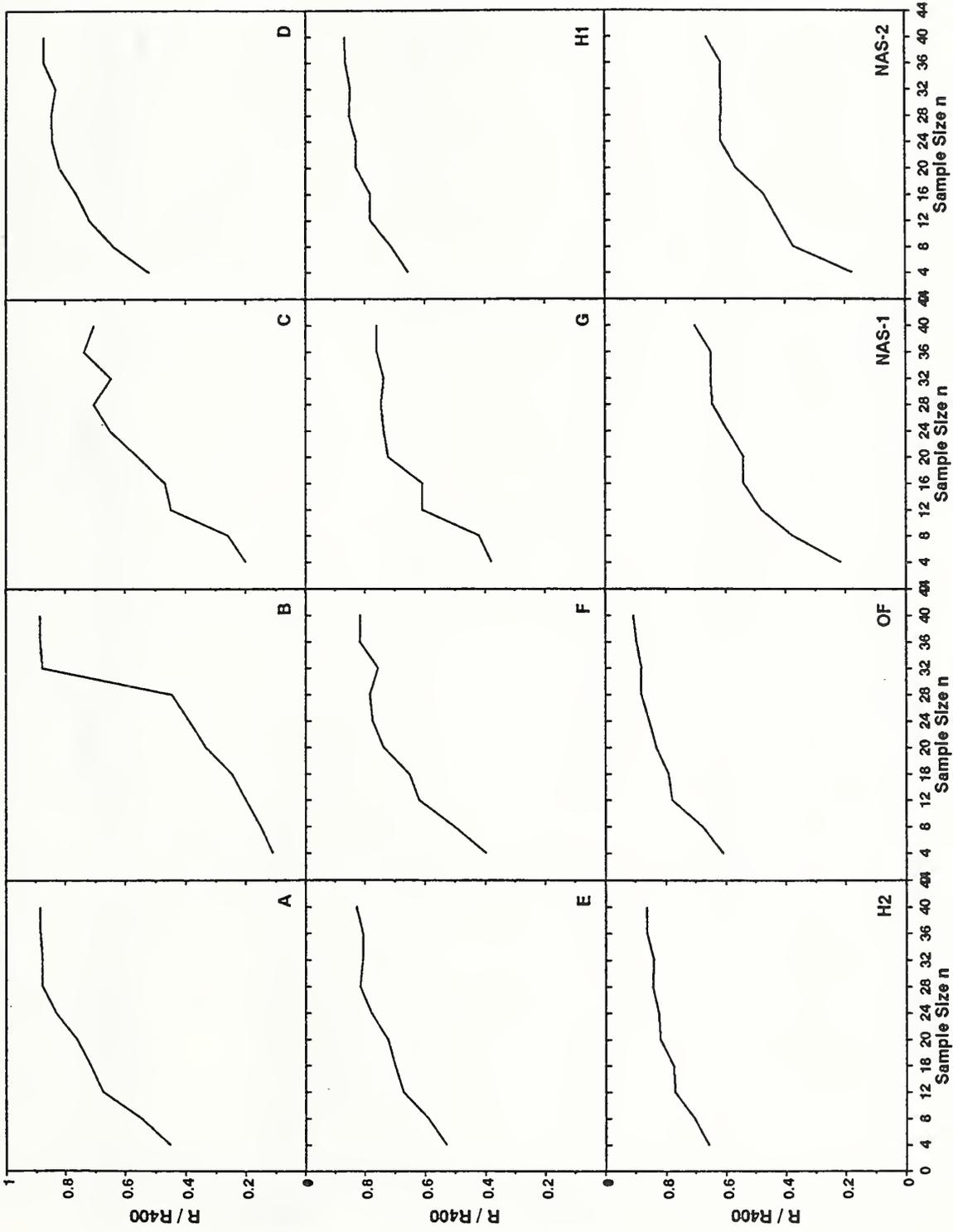


Figure 9-c. First decile of  $R/R400$  as a function of sample size; randomized block design (4 blocks)

# RANDOMIZED BLOCK DESIGN (4 BLOCKS)--MINIMUM PLOTS

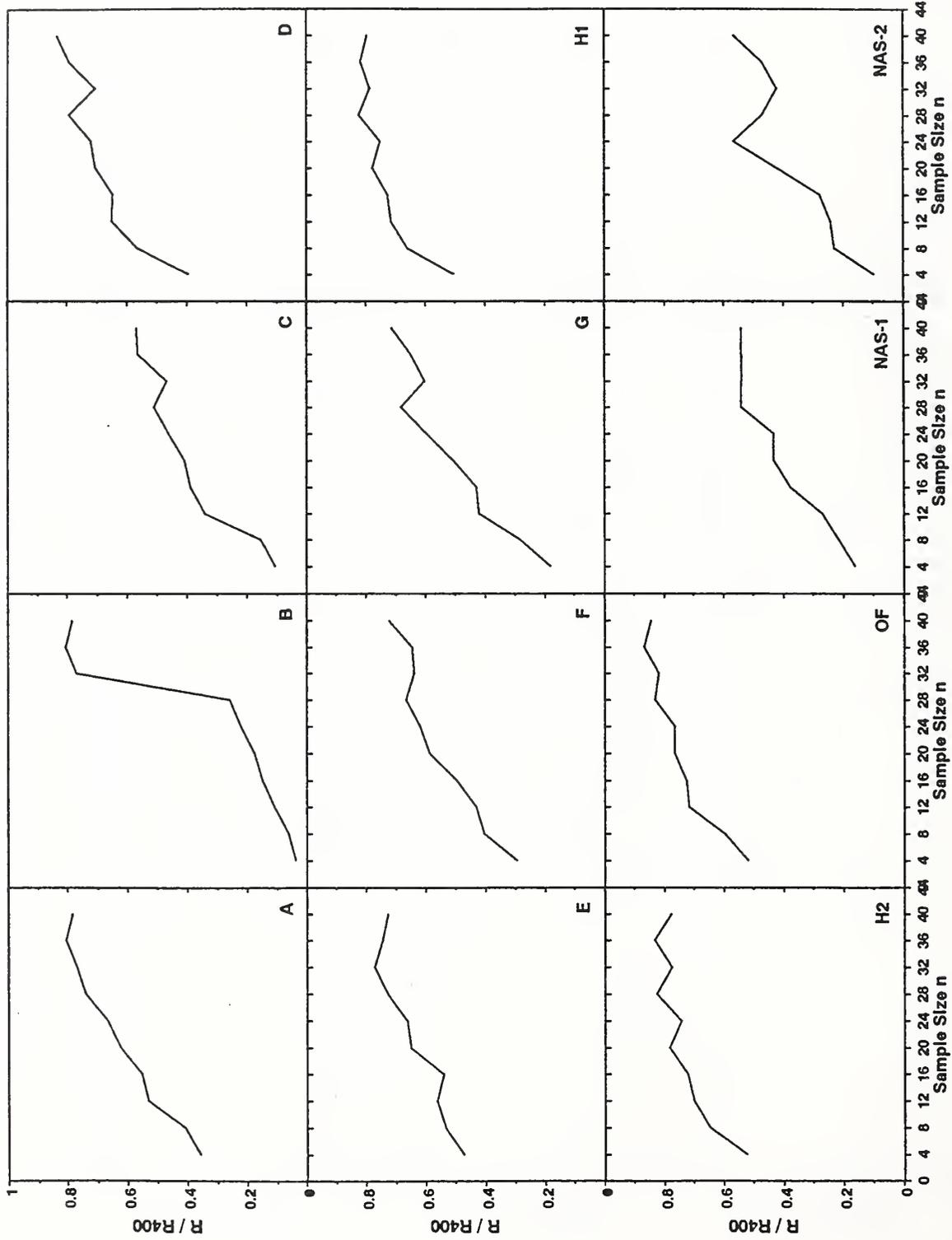
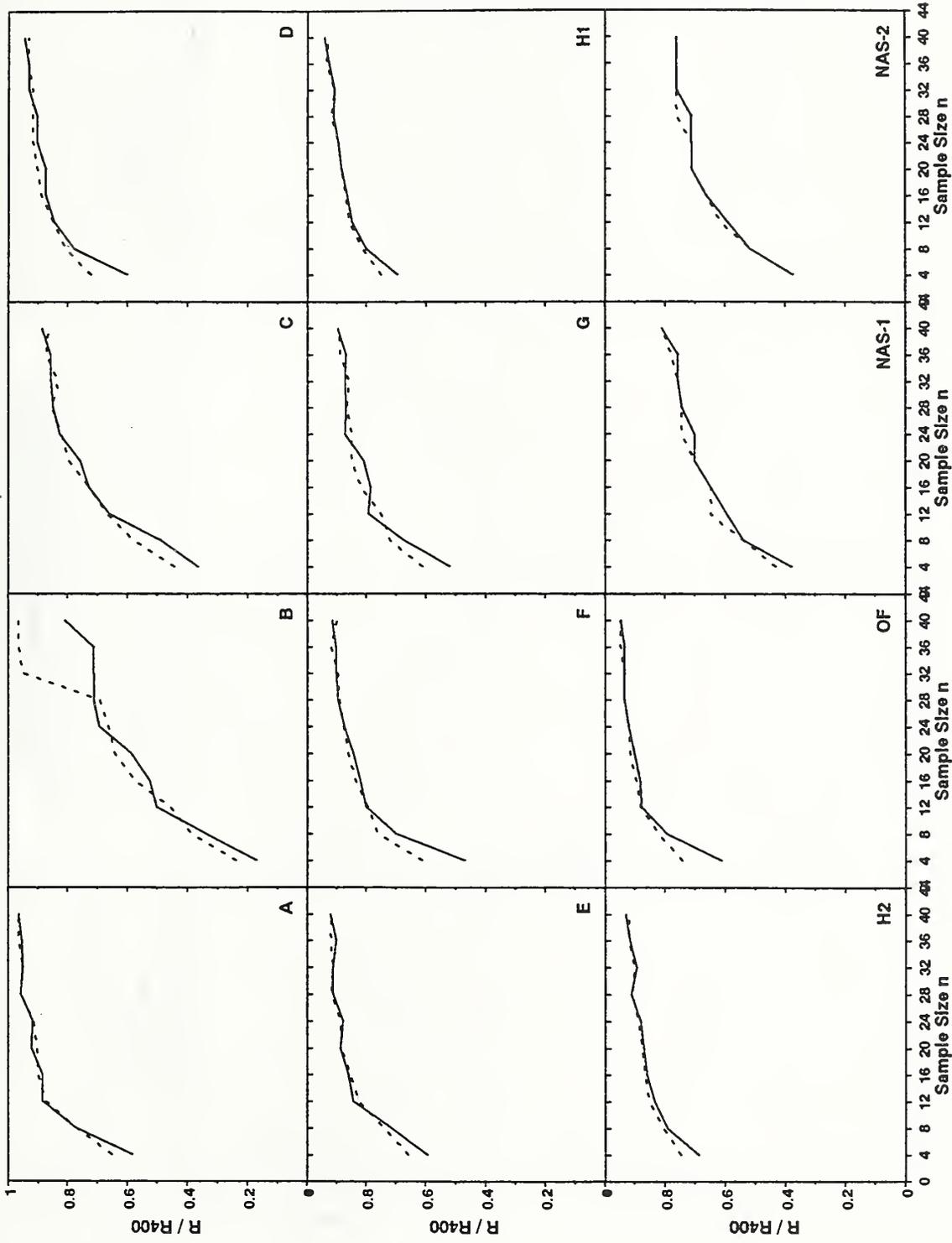


Figure 9-d. Minimum of  $R/R400$  as a function of sample size; randomized block design (4 blocks)

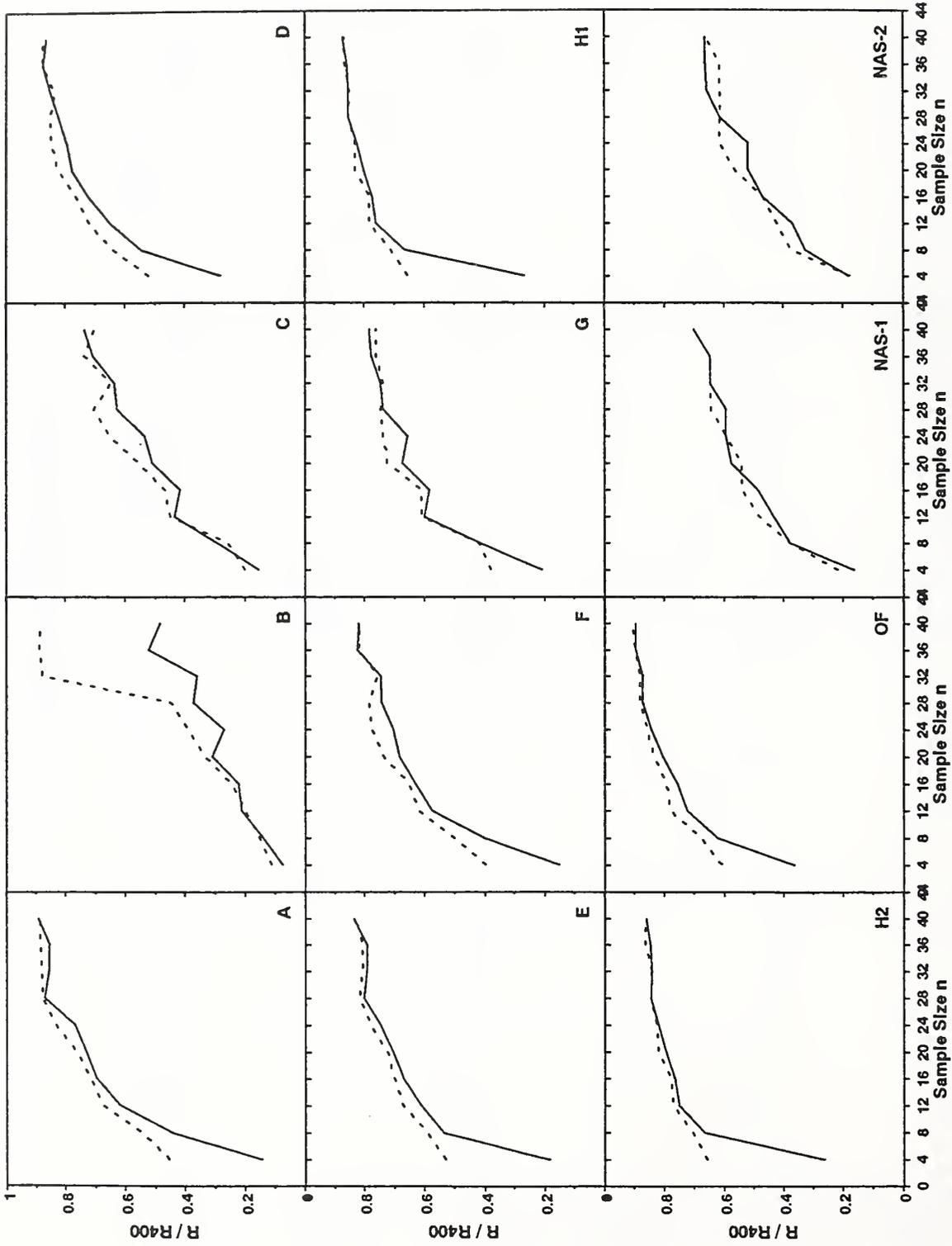
# CRD/RBD(4) COMPARISON--MEDIAN PLOTS



SOLID = COMPLETELY RANDOMIZED DESIGN  
 DOTTED = RANDOMIZED BLOCK DESIGN

Figure 10-a. Comparison of completely randomized design and randomized block designs (4 blocks): medians of  $R/R400$  as a function of sample size

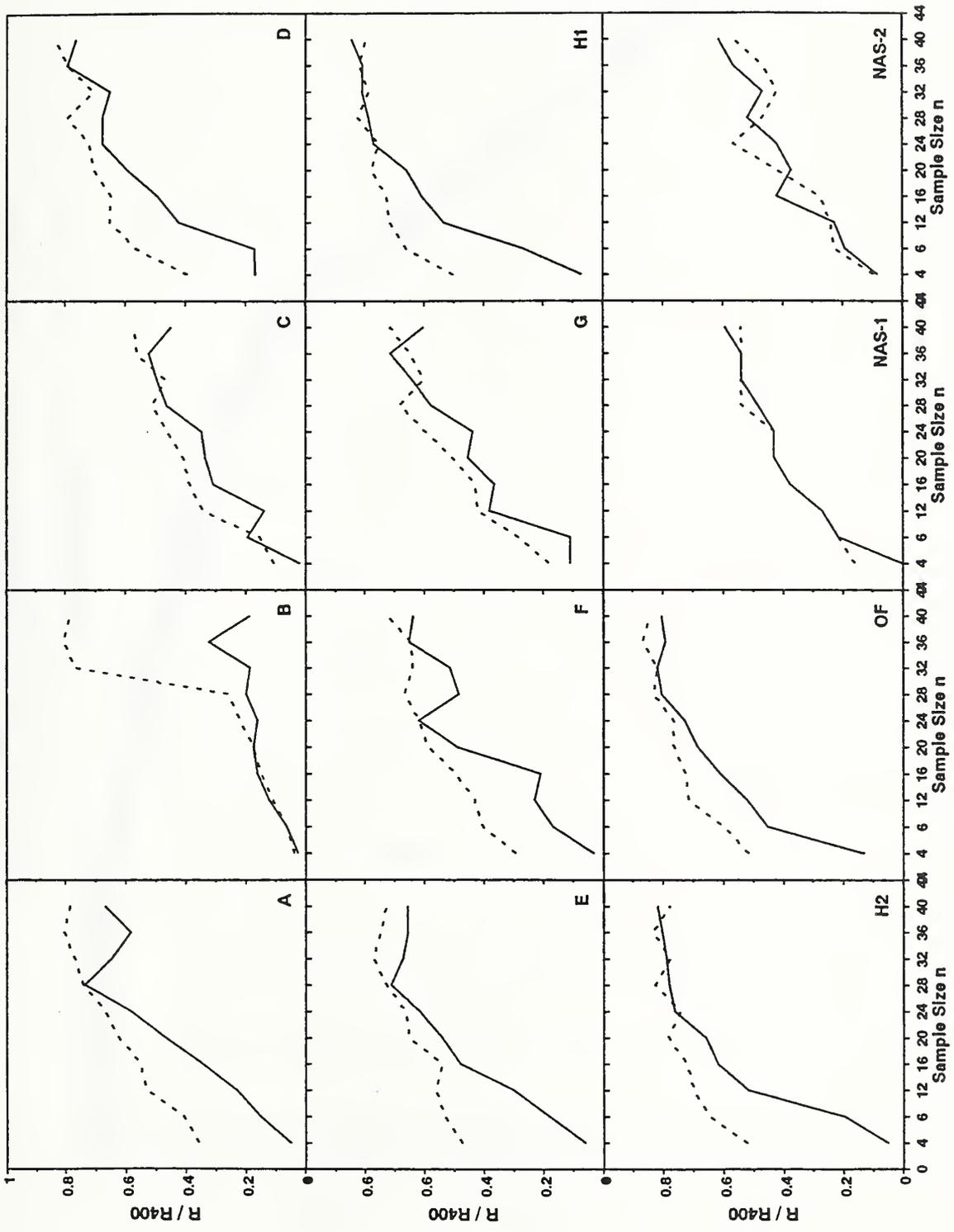
# CRD/RBD(4) COMPARISON--FIRST DECILE PLOTS



SOLID = COMPLETELY RANDOMIZED DESIGN  
 DOTTED = RANDOMIZED BLOCK DESIGN

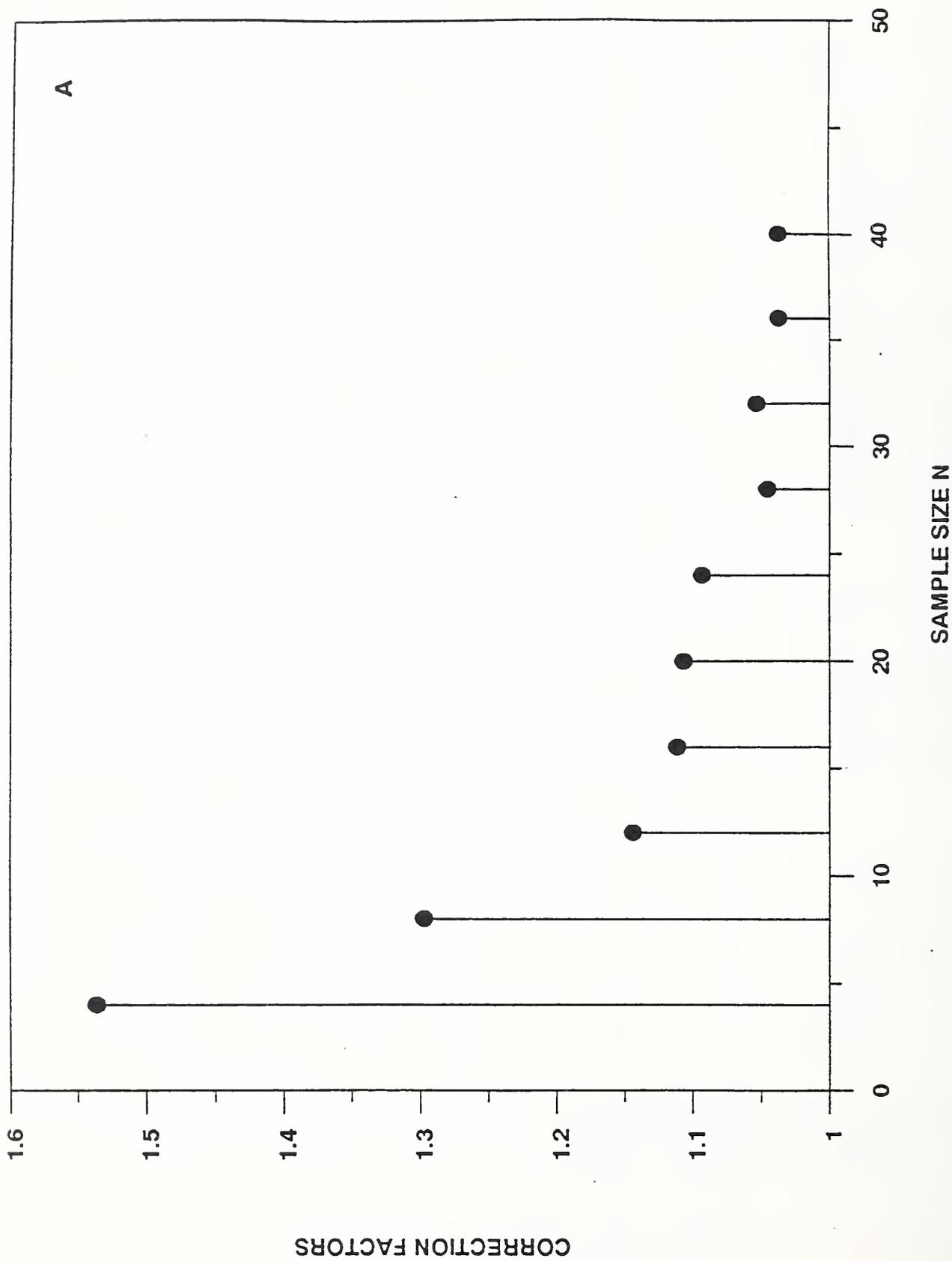
Figure 10-b. Comparison of completely randomized design and randomized block designs (4 blocks): first deciles of R/R400 as a function of sample size

# CRD/RBD(4) COMPARISON--MINIMUM PLOTS



SOLID = COMPLETELY RANDOMIZED DESIGN  
 DOTTED = RANDOMIZED BLOCK DESIGN

Figure 10-c. Comparison of completely randomized design and randomized block designs (4 blocks): minima of  $R/R_{400}$  as a function of sample size



**MEDIAN-UNBIASED CORRECTION FACTORS FOR RANGE**  
 Figure 11. Range bias-correction factors for GE-A;  
 completely randomized design

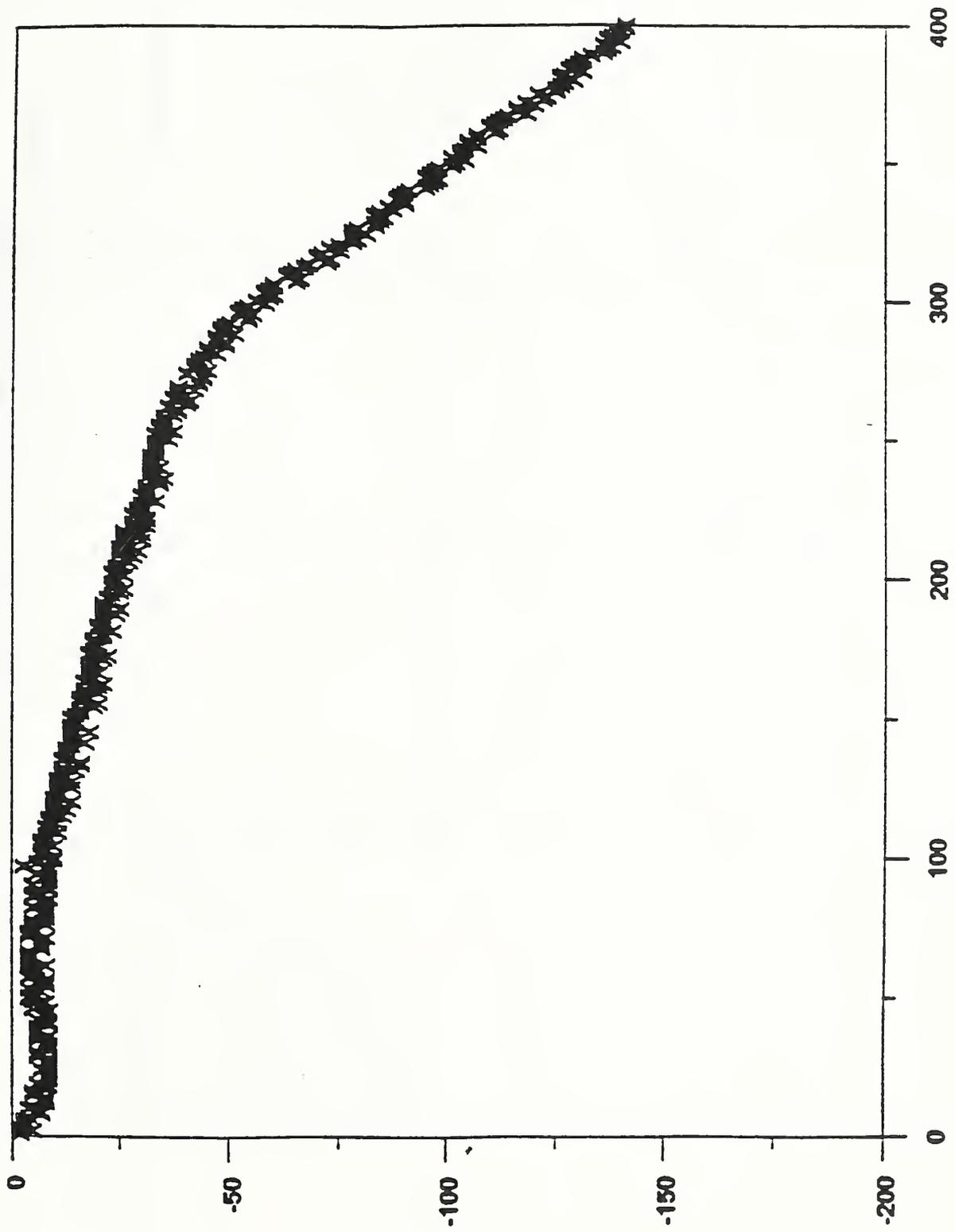


Figure 12. Surface profile for GE-A (400 points)

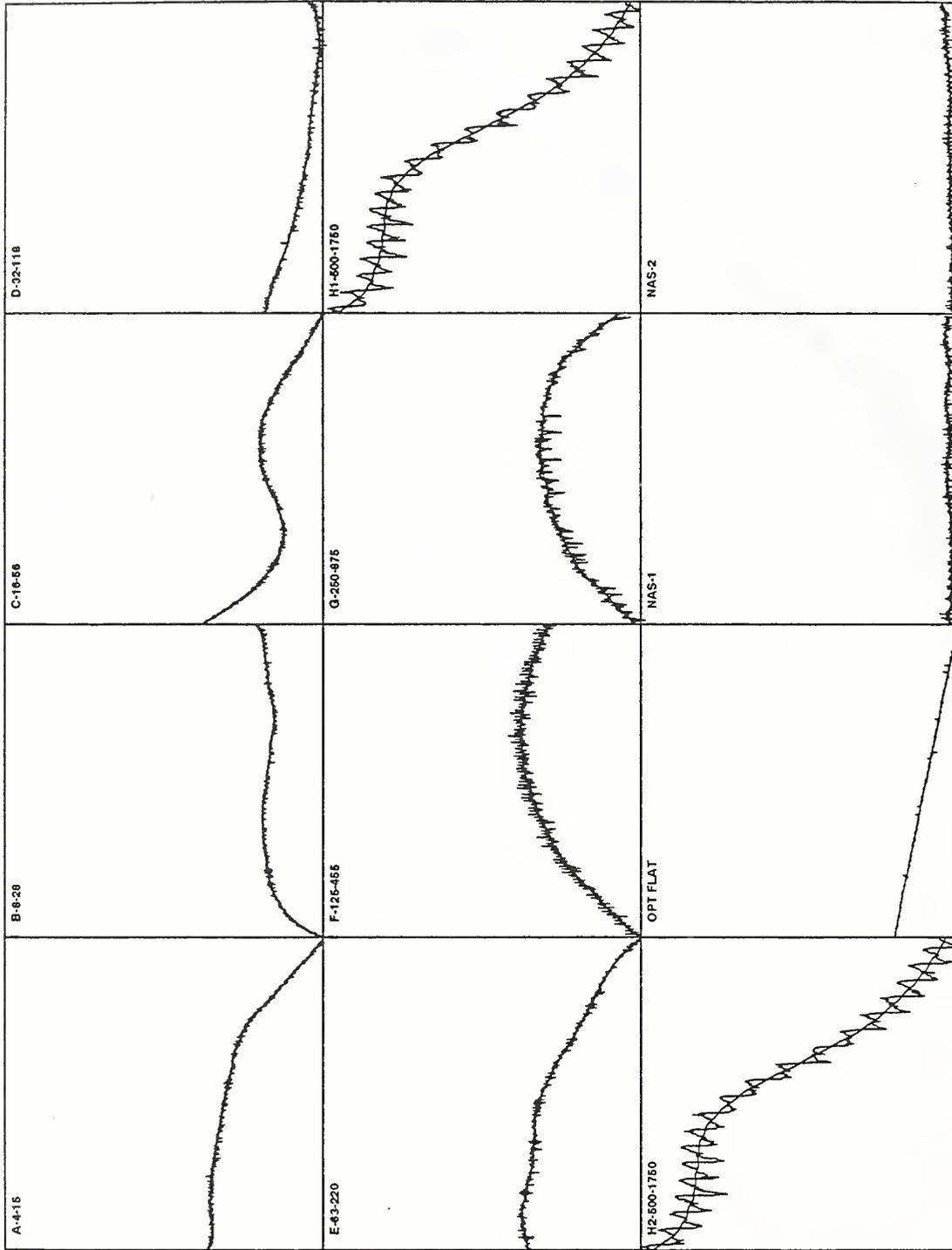


Figure 13. Surface profiles of raw data for all 12 specimens  
(with common scale for all vertical axes)

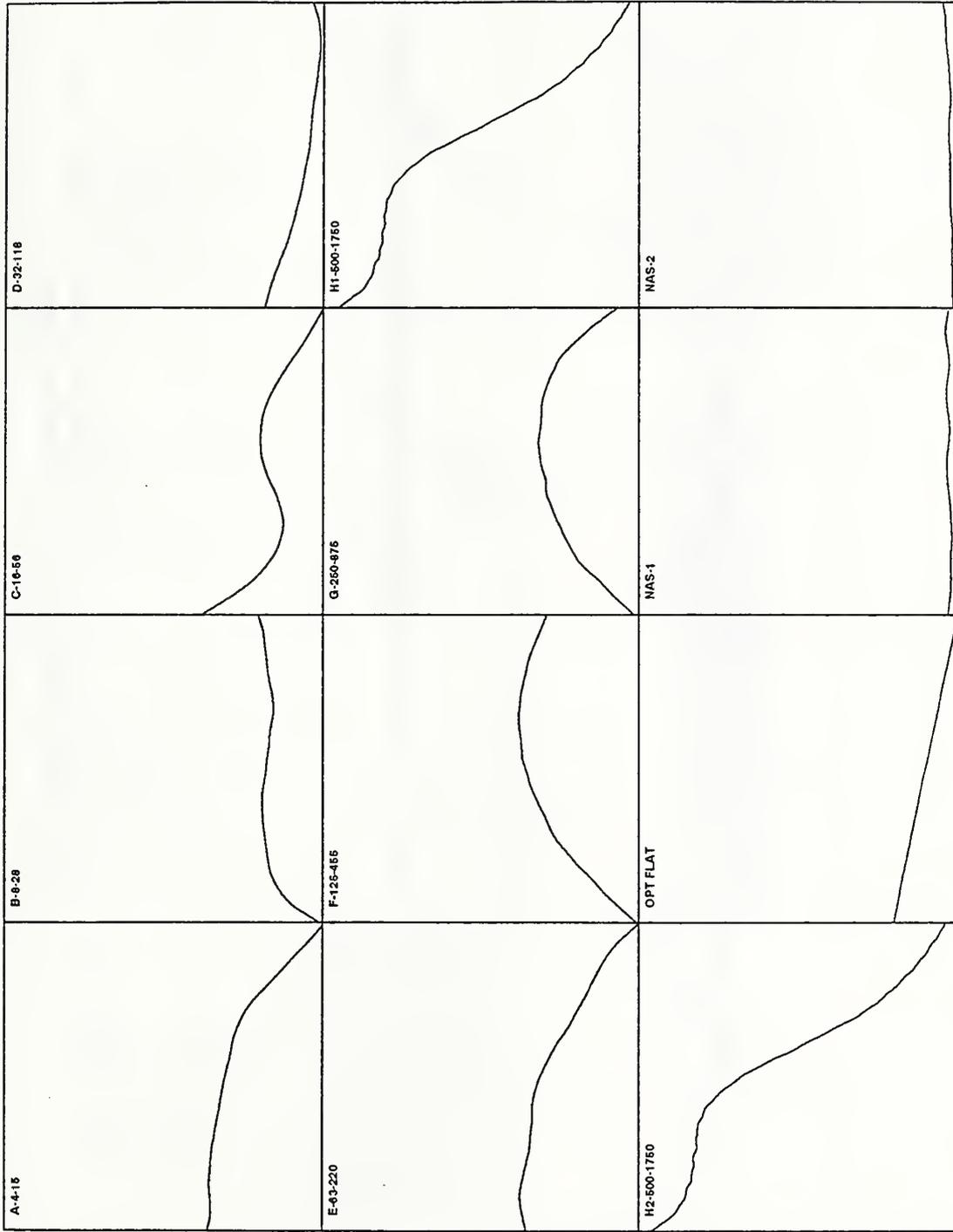


Figure 14. Low-frequency component of the original 12 data sets  
(with common scale for all vertical axes)

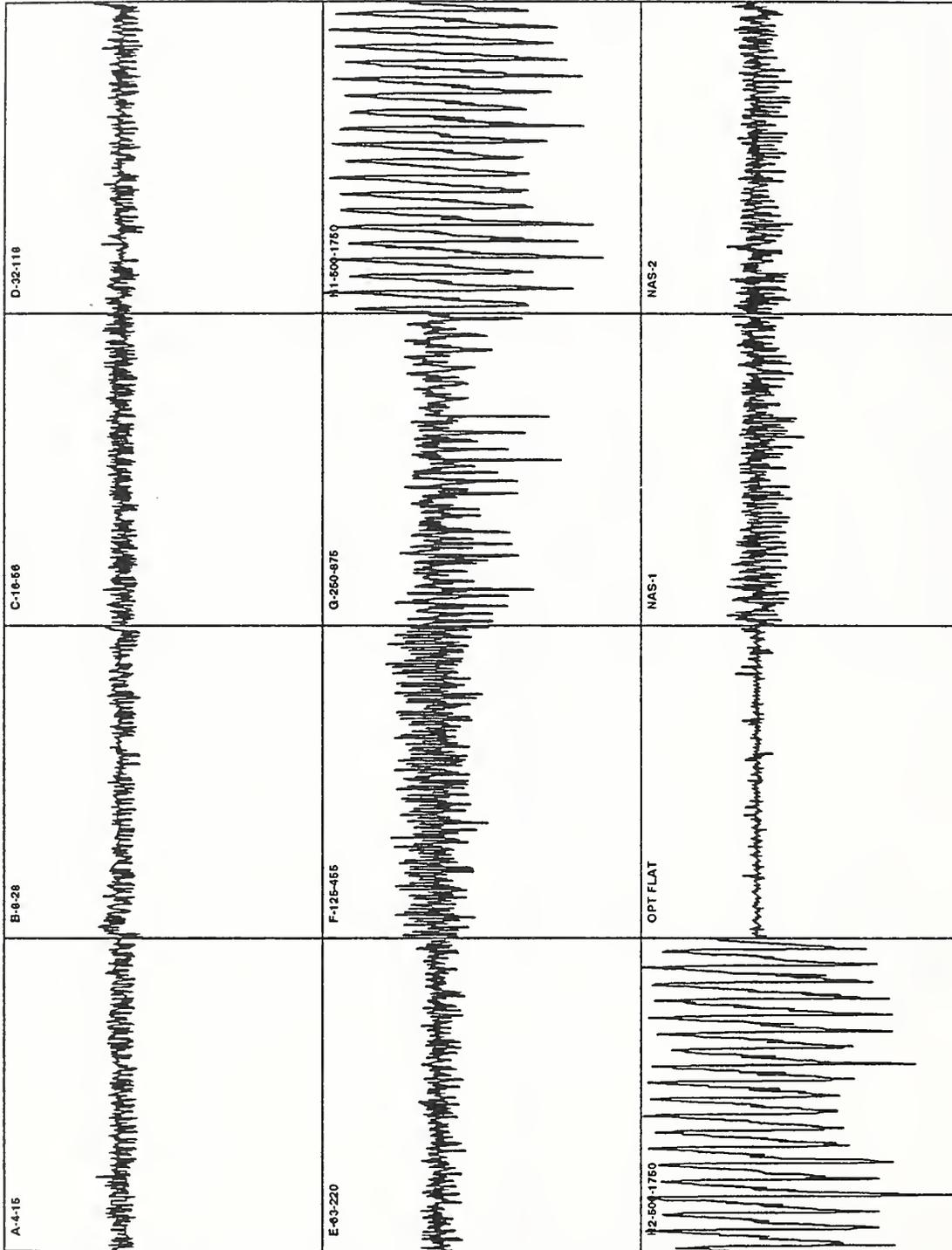


Figure 15. High-frequency component of the original 12 data sets  
(with common scale for all vertical axes)

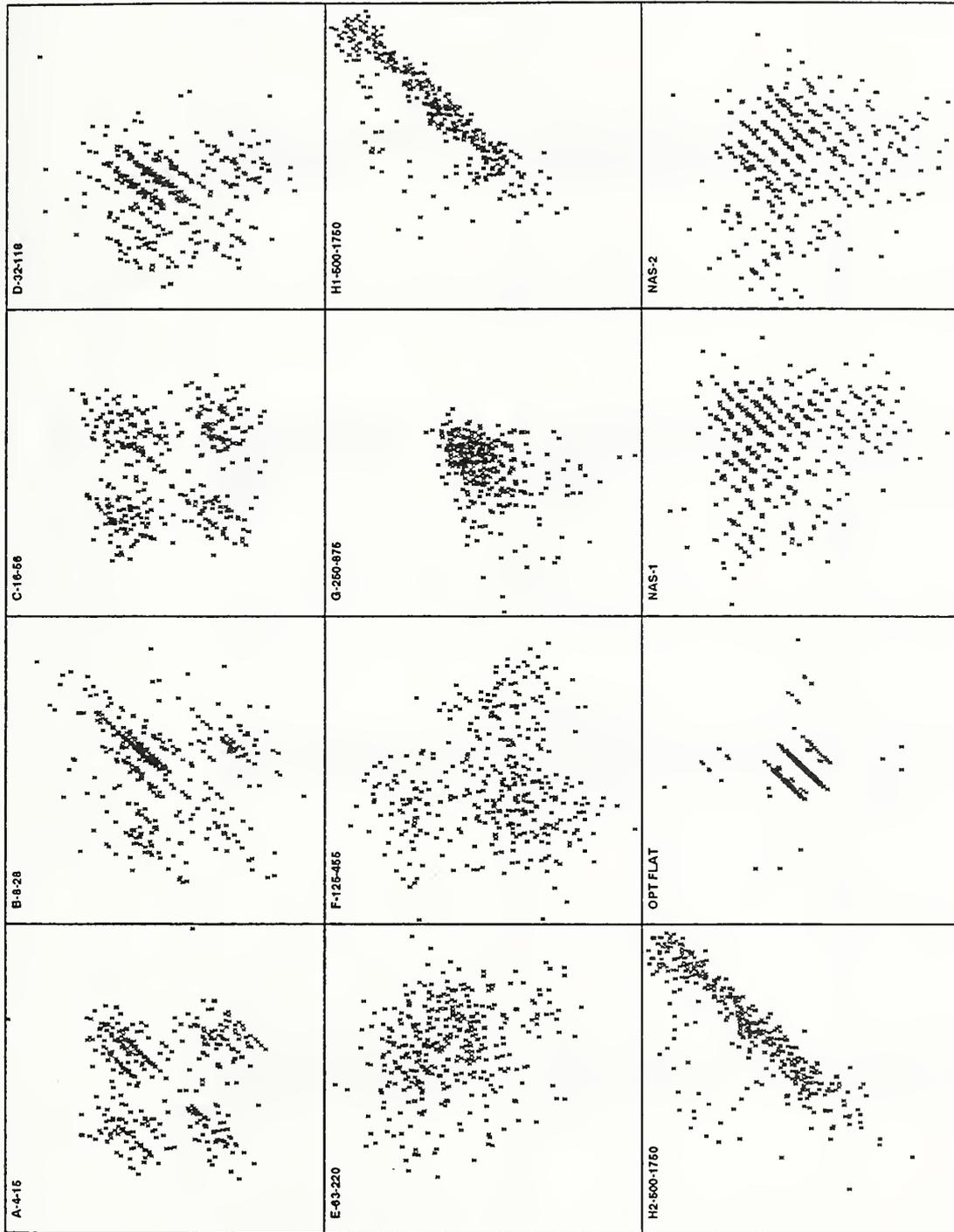


Figure 16. Checking for autocorrelation structure: lag plots ( $y(i)$  versus  $y(i-1)$ ) of the high-frequency component for the original 12 data sets.

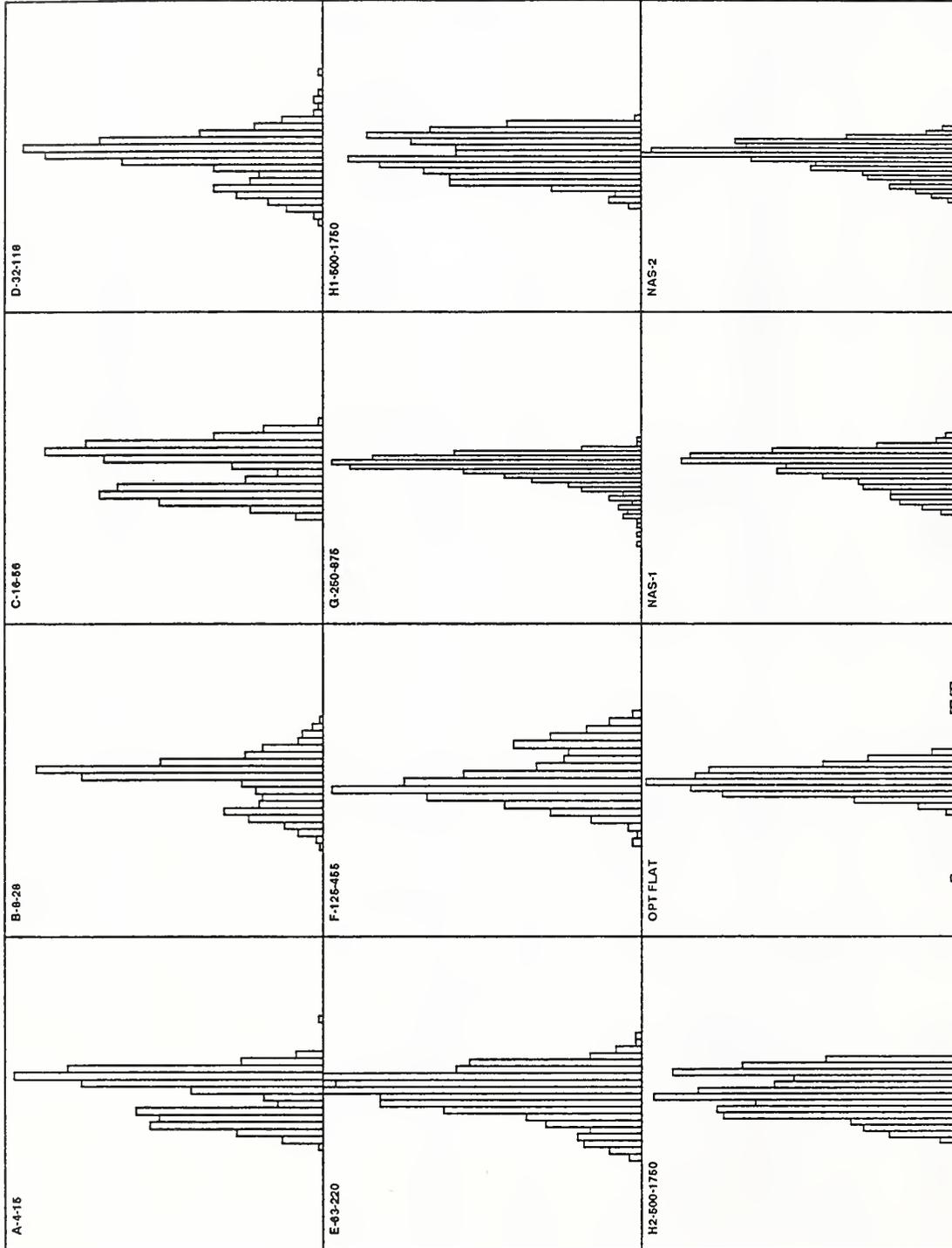


Figure 17. Checking for distributional structure:  
 histograms of the high-frequency component  
 for the original 12 data sets

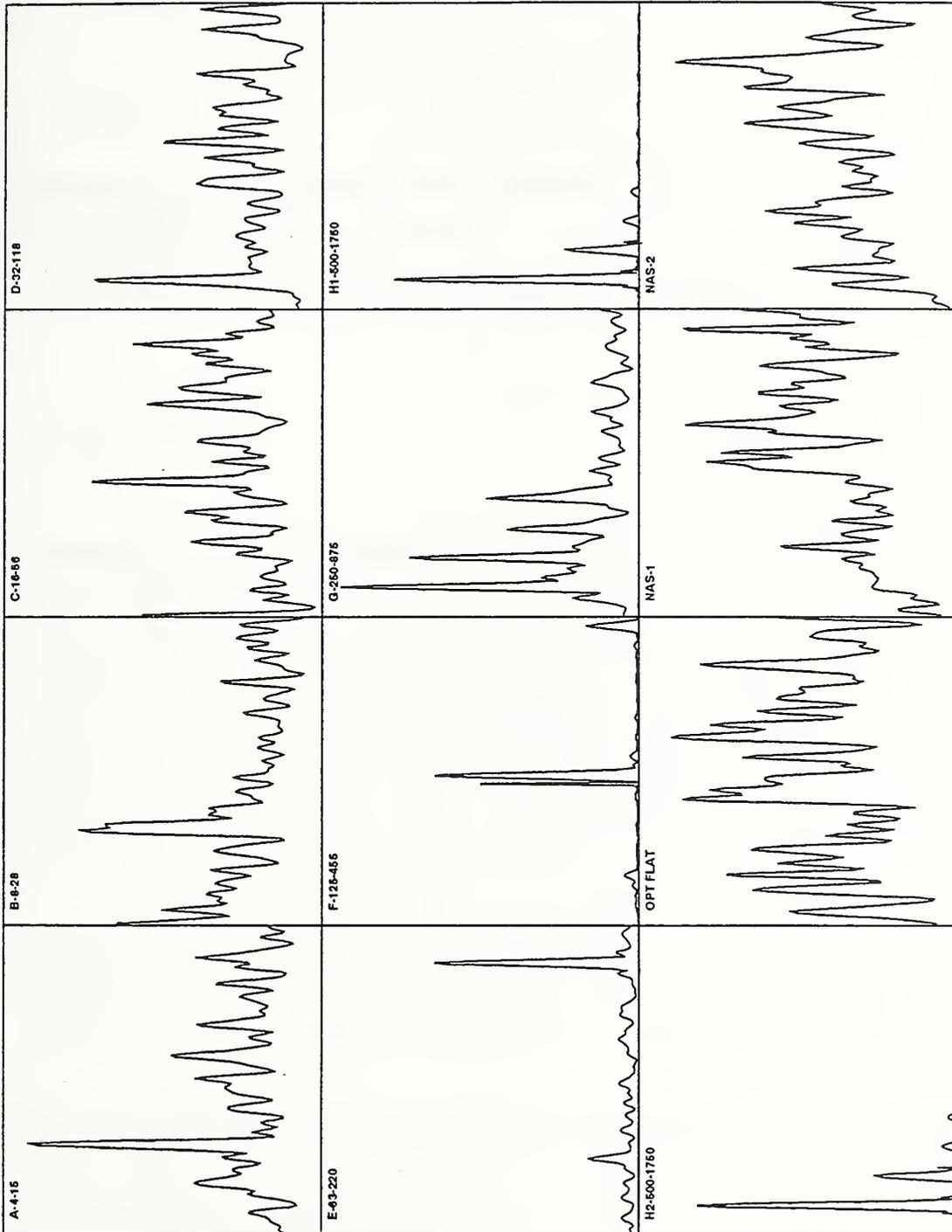


Figure 18. Checking for frequency structure:  
 spectral plots of the high-frequency component  
 for the original 12 data sets



NIST-114A (REV. 3-90)	U.S. DEPARTMENT OF COMMERCE NATIONAL INSTITUTE OF STANDARDS AND TECHNOLOGY	1. PUBLICATION OR REPORT NUMBER NISTIR 4897
<b>BIBLIOGRAPHIC DATA SHEET</b>		2. PERFORMING ORGANIZATION REPORT NUMBER
		3. PUBLICATION DATE AUGUST 1992
4. TITLE AND SUBTITLE  A Comparative Study of CMM Flat-Surface Inspection Plans		
5. AUTHOR(S)  J. J. Filliben, J. P. Hsu, T. W. Hsu, T. H. Hopp		
6. PERFORMING ORGANIZATION (IF JOINT OR OTHER THAN NIST, SEE INSTRUCTIONS) U.S. DEPARTMENT OF COMMERCE NATIONAL INSTITUTE OF STANDARDS AND TECHNOLOGY GAITHERSBURG, MD 20899	7. CONTRACT/GRANT NUMBER	8. TYPE OF REPORT AND PERIOD COVERED
9. SPONSORING ORGANIZATION NAME AND COMPLETE ADDRESS (STREET, CITY, STATE, ZIP)		
10. SUPPLEMENTARY NOTES		
11. ABSTRACT (A 200-WORD OR LESS FACTUAL SUMMARY OF MOST SIGNIFICANT INFORMATION. IF DOCUMENT INCLUDES A SIGNIFICANT BIBLIOGRAPHY OR LITERATURE SURVEY, MENTION IT HERE.)  <p>This document reports on the findings from a collaborative effort by J. P. Hsu of the University of Texas at El Paso, T. W. Hsu (also of the same affiliation), James J. Filliben of the Statistical Engineering Division of National Institute of Standards and Technology (NIST), and Theodore H. Hopp of the Machine Intelligence Group at NIST. The subject of the investigation was inspection sampling plans as related to coordinate measuring machines (CMM). The goal of the investigation was to compare various inspection plans, and--particular-- to make recommendations on minimal sample sizes to achieve pre-specified engineering tolerances on a CMM. The investigation is baseline in the sense that the recommendations flow from a part geometry which is--by design--fundamental (a plane) and a part characteristic which is fundamental (flatness).</p> <p>This report provides details on the engineering considerations that went into setting up and running an appropriate experiment, and the several quantitative and graphical data analysis steps involved in the subsequent statistical analysis. We expect that the information included in this report will be of interest to CMM users, designers, and manufacturers; and we expect such procedural detail to serve as a guideline for other researchers involved on more complicated part geometries and part characteristics.</p>		
12. KEY WORDS (6 TO 12 ENTRIES; ALPHABETICAL ORDER; CAPITALIZE ONLY PROPER NAMES; AND SEPARATE KEY WORDS BY SEMICOLONS)  CMM's; coordinate measuring machines; flatness; inspection plans; sampling plans; sample size determination		
13. AVAILABILITY <input checked="" type="checkbox"/> UNLIMITED FOR OFFICIAL DISTRIBUTION. DO NOT RELEASE TO NATIONAL TECHNICAL INFORMATION SERVICE (NTIS). ORDER FROM SUPERINTENDENT OF DOCUMENTS, U.S. GOVERNMENT PRINTING OFFICE, WASHINGTON, DC 20402. <input checked="" type="checkbox"/> ORDER FROM NATIONAL TECHNICAL INFORMATION SERVICE (NTIS), SPRINGFIELD, VA 22161.	14. NUMBER OF PRINTED PAGES 52	15. PRICE A04

ELECTRONIC FORM





